

国际机械工程先进技术译丛

产品可靠性、维修性及保障性手册

(原书第2版)

Product Reliability, Maintainability, and Supportability Handbook

Second Edition

(美) Michael Pecht 著

王军锋 陈云斌 周宪 等译





机械工业出版社
CHINA MACHINE PRESS

国际机械工程先进技术译丛

产品可靠性、维修性及保障性手册

（原书第2版）

Product Reliability, Maintainability, and Supportability Handbook

(Second Edition)

（美） Michael Pecht 著

王军锋 陈云斌 周 宪 等译

闫镜欣 唐璐敏 审校



机 械 工 业 出 版 社

译 丛 序 言

一、制造技术长盛永恒

先进制造技术是20世纪80年代提出的，它由机械制造技术发展而来，通常可以认为它是将机械、电子、信息、材料、能源和管理等方面的技术，进行交叉、融合和集成，综合应用于产品全生命周期的制造全过程，包括市场需求、产品设计、工艺设计、加工装配、检测、销售、使用、维修、报废处理、回收利用等，以实现优质、敏捷、高效、低耗、清洁生产，快速响应市场的需求。因此，当前的先进制造技术是以产品为中心，以光机电一体化的机械制造技术为主体，以广义制造为手段，具有先进性和时代感。

制造技术是一个永恒的主题，与社会发展密切相关，是设想、概念、科学技术物化的基础和手段，是所有工业的支柱，是国家经济与国防实力的体现，是国家工业化的关键。现代制造技术是当前世界各国研究和发展的主题，特别是在市场经济高度发展的今天，它更占有十分重要的地位。

信息技术的发展并引入到制造技术，使制造技术产生了革命性的变化，出现了制造系统和制造科学。制造系统由物质流、能量流和信息流组成，物质流是本质，能量流是动力，信息流是控制；制造技术与系统论、方法论、信息论、控制论和协同论相结合就形成了新的制造学科。

制造技术的覆盖面极广，涉及机械、电子、计算机、冶金、建筑、水利、电子、运载、农业以及化学、物理学、材料学、管理科学等领域。各个行业都需要制造业的支持，制造技术既有普遍性、基础性的一面，又有特殊性、专业性的一面；制造技术既有共性，又有个性。

我国的制造业涉及以下三个方面的领域：

- 机械、电子制造业，包括机床、专用设备、交通运输工具、机械设备、电子通信设备、仪器等；
- 资源加工工业，包括石油化工、化学纤维、橡胶、塑料等；
- 轻纺工业，包括服装、纺织、皮革、印刷等。

目前世界先进制造技术沿着全球化、绿色化、高技术化、信息化、个性化和服务化、集群化六个方向发展，在加工技术上主要有超精密加工技术、纳米加工技术、数控加工技术、极限加工技术、绿色加工技术等，在制造模式上主要有自动化、集成化、柔性化、敏捷化、虚拟化、网络化、智能化、协作化和绿色化等。

二、图书交流源远流长

近年来，国际间的交流与合作对制造业领域的发展、技术进步及重大关键技术

的突破起到了积极的促进作用，制造业科技人员需要及时了解国外相关技术领域的最新发展状况、成果取得情况及先进技术应用情况等。

必须看到，我国制造业与工业发达国家相比，仍存在较大差距。因此必须加强原始创新，在实践中继承和创新，学习国外的先进制造技术和经验、引进消化吸收创新，提高自主创新能力，形成自己的创新体系。

国家、地区间的学术、技术交流已有很长的历史，可以追溯到唐朝甚至更远一些，唐玄奘去印度取经可以说是一次典型的图书交流佳话。图书资料是一种传统、永恒、有效的学术、技术交流方式，早在20世纪初期，我国清代学者严复就翻译了英国学者赫胥黎所著的《天演论》，其后学者周建人翻译了英国学者达尔文所著的《物种起源》，对我国自然科学的发展起到了很大的推动作用。

图书是一种信息载体，图书是一个海洋，虽然现在已有网络、光盘、计算机等信息传输和储存手段，但图书更具有广泛性、适应性、系统性、持久性和经济性，看书总比在计算机上看资料要方便习惯，不同层次的要求可以参考不同层次的图书，不同职业的人员可以参考不同类型的技术图书，同时它具有比较长期的参考价值和收藏价值。当然，技术图书的交流具有时间上的滞后性，不够及时，翻译的质量也是个关键问题，需要及时、快速、高质量的出版工作支持。

机械工业出版社希望能够在先进制造技术的引进、消化、吸收、创新方面为广大读者作出贡献，为我国的制造业科技人员引进、纳新国外先进制造技术的出版资源，翻译出版国际上优秀的制造业先进技术著作，从而能够提升我国制造业的自主创新能力，引导和推进科研与实践水平的不断进步。

三、选译严谨质高面广

1）精品重点高质　本套丛书作为我社的精品重点书，在内容、编辑、装帧设计等方面追求高质量，力求为读者奉献一套高品质的丛书。

2）专家选译把关　本套丛书的选书、翻译工作均由国内相关专业的专家、教授、工程技术人员承担，充分保证了内容的先进性、适用性和翻译质量。

3）引纳地区广泛　主要从制造业比较发达的国家引进一系列先进制造技术图书，组成一套《国际机械工程先进技术译丛》。当然其他国家的优秀制造科技图书也在选择之内。

4）内容先进丰富　在内容上应具有先进性、经典性、广泛性，应能代表相关专业的技术前沿，对生产实践有较强的指导、借鉴作用。本套丛书尽量涵盖制造业各行业，例如机械、材料、能源等，既包括对传统技术的改进，又包括新的设计方法、制造工艺等技术。

5）读者层次面广　面对的读者对象主要是制造业企业、科研院所的专家、研究人员和工程技术人员，高等院校的教师和学生，可以按照不同层次和水平要求各取所需。

四、衷心感谢不吝指教

首先要感谢许多积极热心支持出版《国际机械工程先进技术译丛》的专家学者，积极推荐国外相关优秀图书，仔细评审外文原版书，推荐评审和翻译的知名专家，特别要感谢承担翻译工作的译者，对各位专家学者所付出的辛勤劳动表示深切敬意，同时要感谢国外各家出版社版权工作人员的热心支持。

本套丛书希望能对广大读者的工作提供切实的帮助，欢迎广大读者不吝指教，提出宝贵意见和建议。

机械工业出版社

序　言

企业必须在产品开发过程中实施一些措施，以确保产品的可靠性。这些措施会通过零部件（材料）选择、产品设计、制造、装配、运输、装卸、运行、维护和修理等过程影响可靠性。本书将介绍以下内容：

1）根据目标生命周期应用条件和预期产品性能等因素，制定切实可行的产品可靠性要求。产品的可靠性要求必须考虑客户需求，以及制造商满足这些需求的能力。

2）通过评估相关制造、装配、存储、装卸、运输、运行和维修条件定义产品生命周期条件。

3）确保供应链的参与者有能力提供符合要求的零部件（材料），并能提供满足最终可靠性目标所需的服务。

4）选择有质量保证的零部件（材料），这些零部件要能在应用中实现预期性能和可靠性的要求。

5）确定可能失效产品的潜在失效模式、失效位置及失效机理。

6）设计工艺能力（也就是制造和装配中可控制的质量级别），并考虑潜在失效模式、失效位置，以及从失效的物理学分析和生命周期轮廓中获取的失效机理。

7）赋予产品合格的质量，以在预期生命周期条件下验证其可靠性。质量检测包括的活动要能保证标称设计和制造规范将满足或超过可靠性目标。

8）确定所有制造和装配过程生产的产品是否在设计所要求的统计过程窗内。材料特性和制造工艺的变异将影响产品的可靠性。因此，整个过程必须是经过鉴定、可度量以及可监控的。

9）利用闭合循环式的监测来管理生命周期中产品的使用。

第1章：产品效能与价值。此章给出了产品效能的定义，并讨论了产品效能及相关函数（可用性、可信性和能力）之间的关系，最后还讨论了责任分配和产品价值。

第2章：与可靠性相关的概念。此章阐述了可靠性的数学理论基础，主要内容是可靠性和不可靠性函数、概率密度函数、故障率、条件可靠性函数和关键失效时间。

第3章：统计推论概念。此章引入统计推论概念，用它们来分析从观测数据得到的概率模型，主要讨论了统计估计、假设检验的基本类型和可靠性回归模型拟合。

第4章：产品可靠性分析的实用概率分布。此章引入了两种基本概率分布类型：离散和连续概率分布，讨论了在产品可靠性建模和故障率估计中常用的两种离散分布（二项分布和Poisson分布）和四种连续型分布（Weibull分布、指数分布、正态分布、对数正态分布）。

第5章：置信区间。此章给出了置信区间的概念，讨论了它和容差、样本大小以及置信水平的关系，并提供了计算估计置信区间的实例。

第 6 章：硬件可靠性。此章通过失效模型和相关案例，主要讨论可靠性评估，以及相关的用于工程硬件的试验技术。以微电子封装中焊接连接为例，阐述了失效物理的概率方法在可靠性预测和建模过程中的应用。

第 7 章：软件可靠性。此章给出了软件、软件可靠性、软件质量和软件安全的定义，还讨论了用以改进并评估软件可靠性的软件开发模型和技术。

第 8 章：失效模式、机理及影响分析。对于可靠性产品的开发过程来说，导致产品失效的失效机理知识是实现合理设计、开发可靠性产品的关键。此章提出了一个称为失效模式、机理及影响分析（FMMEA）的新方法，用以确定潜在失效机理以及潜在失效模式的失效模型，并将失效机理进行优先度排序。FMMEA 提升了失效模型及影响分析（FMEA）的价值，失效模式、影响及危害性分析（FMECA）通过确定“高优先度失效机理”帮助相关人员制定策略，以减轻失效影响。在 FMMEA 中发现的失效机理原因和影响，有助于提升产品开发的效率和经济效益。

第 9 章：可靠性设计。要让产品达到可靠性目标，有一些步骤是必不可少的。此章大致介绍了产品需求和约束、产品生命周期条件、零件选择和管理、失效模式、失效机理、失效影响分析、设计技术、质量检测、制造和装配以及闭合循环监测等方面的内容。

第 10 章：系统可靠性建模。此章介绍了如何把来自于零部件和子系统的可靠性信息结合到系统的可靠性计算中。可靠性框图可用于表达系统逻辑结构，并为系统开发可靠性模型。同时，本章还介绍了用于系统可靠性建模的故障树分析。

第 11 章：冗余和容错产品的可靠性分析。设计有容错能力产品的目的在于让产品在某些零部件失效的情况下持续正常运行。此章介绍了用户在各种容错条件下评估可靠性的方法。

第 12 章：可维修产品的可靠性模型和数据分析。可维修产品通常会出现磨损，此章介绍了用于可维修产品（特别是非电子产品）失效的建模和分析方法，提供了描述可维修产品可靠性行为的分析背景和数据分析技术。

第 13 章：持续的可靠性改进。可靠性改进技术可应用于：通过了主要硬件和软件设计审查的新产品；制造商想让其更具竞争力的已开发产品；没有满足客户对可靠性表现需求的已有产品。此章讨论了持续改进项目的可靠性增长、加速测试和管理的原则。

第 14 章：后勤保障。应用到产品的综合后勤保障构成了维修和保障的生命周期方法。此章讨论了可靠性对于后勤保障需求的影响，着重阐述了产品、装备或组件的可靠性是如何影响备用或维修零件、保障设备和维修人员的。

第 15 章：产品效能和成本分析。此章阐述了如何把可靠性、维修性数据与产品性能结合起来，最终评价总体产品效能；如何引入成本概念，为设计决策提供更完整的支持。

第 16 章：工艺能力与过程控制。质量是衡量产品满足制造商工艺标准要求程度的量度。此章介绍了加工能力的概念和过程控制的基本统计技术。通过一些实例，介绍了平均抽检质量、加工能力、缺陷计算和统计过程控制的概念。

译　者　序

企业在产品的整个生命周期中，所实施的策略和所展开的活动会直接影响到产品的可靠性。高度可靠的产品能防止故障和事故，尤其是避免灾难性的事故发生；其生命周期费用较低，停机时间少，利用率高；进而能够提高企业信誉，增强竞争力，扩大市场份额，增加经济效益。如何提高产品的可靠性已经成为现代企业发展不可回避的问题。

可靠性是产品实现其既定功能的必要前提，但完全可靠的产品是不存在的。产品在其使用寿命内，肯定会因各种各样的问题而发生失效，出现故障。产品的维修以及平时的例行维护的难易程度和所要付出的代价会因制造工艺、产品结构、运行环境等因素而有所差异，这也就是产品的维修性。良好的后勤保障能为维修和维护提供充足的检测设备、产品备件和维修设备等，使得产品更易于维修和维护。这些正是本书所要阐述的内容。

《产品可靠性、维修性及保障性手册》的第1版出版于1995年。时隔15年之后，科学技术的发展为我们带来了更多、更新的产品，同时也带来了产品可靠性、维修性和保障性方面的新问题。原书的编者和各章节的作者根据他们在此期间的研究成果，共同出版了该书的第2版。除了第1版中涵盖的可靠性分析的数学理论基础、软件和硬件的可靠性分析、可靠性增长过程、加速寿命试验等内容之外，此版本新增加的内容包括：失效模式、机理和影响分析（FMMEA）；可靠性指标的置信区间及其与产品质量保证措施之间的关系；过程控制和过程能力，以及它们与产品可靠性的关系；冗余系统的可靠性。

本书的主编是 Michael Pecht，他不仅是 IEEE 的资深会士（IEEE Fellow），还是 ASME 和 IMAPS 的会士。他曾获得 IEEE 可靠性协会终生成就奖、欧洲微纳米可靠性奖、3M 研究奖、William D. Ashman 成就纪念奖。同时，他还曾兼任 *IEEE Transactions on Reliability* 期刊的主编和 *IEEE Spectrum* 期刊的专家顾问，现分别任国际期刊 *Microelectronics Reliability* 和 *IEEE Transactions on Components and Packaging Technology* 的主编和副主编。Pecht 编著了诸如 *Handbook of Electronic Package Design*、*Handbook of Electronic Package Design* 和 *Parts Selection and Management* 等关于电子产品开发、使用和供应链管理等方面的书籍20余本，发表学术论文400余篇。

本书各章节的作者大多是在国防、军事等国家部门的可靠性领域从业多年的工程技术人员。例如 Harold S. Balaban 任职于美国国防分析研究所，Criscimagna 咨询有限公司的总裁 Ned Criscimagna 是可靠性咨询业务的资深专家，中国深圳微软亚洲硬件研发中心主管 Jun Ming Hu 致力于提高企业产品的可靠性。此外，还有多年从事可靠性研究的研究人员，如马里兰大学工程研究中心可靠性项目主管 David Weiss

和 ARINC 研究公司的首席工程师 Robert M. Hecht 等。他们都已在产品可靠性、维修性和保障性的工程应用和研究方面取得了显著成就，本书正是他们多年研究和实践成果的汇总。

因此，本书可谓是可靠性、维修性和保障性技术领域的权威指南。正是由于本书主编和作者丰富的从业经验和研究经历，才使得本书成为一本可靠性方面的优秀畅销书籍，一本对于可靠性工程人员和学习相关知识和技术的人们来说不可缺少的手册。感谢本书的主编 Michael Pecht 教授和各章节的作者们，正是他们的不懈努力和无私奉献，我们才能看到这本巨著。

本书翻译和审校工作的大致分工如下：西南科技大学的王军锋翻译了第 1、2、5～10 章，并负责全书的统稿工作；西安工业大学的周宪翻译了第 11、12 章；通用电气中国研发中心的陈云斌翻译了第 13、15、16 章；西北工业大学的石真真翻译了第 3、4 章；西北工业大学的王宁翻译了第 14 章。锐德世系统设备有限公司的高级质量和可靠性工程师唐璐敏审阅了本书的初稿，西北工业大学的闫锋欣完成了各章的主审和全书的终审工作。其他参与本书部分翻译工作的人员还有：苏力争、杨振朝、张莹、曹少飞、戚彬、陈伟鸿、谭邦建、饶锦锋、张延超、王淑侠、王振军、顾婷、李妙玲、徐景辉、戴高明和周杰。

感谢机械工业出版社的编辑们，正是他们的辛苦工作才使本书的中文版得以问世；感谢中国可靠性论坛（www. kekaoxing. com/club/）的版主 Fanweipin 和 Cliffcrag，正是他们促成了本书的译者团队，也感谢该论坛所提供的大量参考资料；感谢戚彬、王淼、刘晓霞和韩卫荣为我分担了不少日常工作，使我能够潜心于本书的翻译工作；感谢我的妻子郭偎和女儿王若水，她们是我努力工作的动力源泉；最后，感谢所有为本书的翻译出版工作提供了帮助的人们！

在翻译的过程中，译者已及时对原书中出现的些许印刷错误做了订正，在此不再一一指出。

鉴于译者的英文理解能力和中文表达能力有限，译文难免出现错误和纰漏，希望领域内的各位同行和专家予以批评指正。

衷心地希望本书能对您的工作和学习有所帮助。

王军锋

主 编 介 绍

Michael Pecht 是香港城市大学电子工程系的客座教授，电子工程硕士，美国威斯康星大学麦迪逊分校机械工程的硕士和博士。现为专业工程师、IEEE 会士（IEEE Fellow）、ASME 会士（ASME Fellow）和 IMAPS 会士（IMAPS Fellow）。2008 年，获得可靠性领域的最高奖项——IEEE 可靠性协会终生成就奖。曾任 *IEEE Transactions on Reliability* 主编八年，*IEEE Spectrum* 咨询顾问。现为 *Microelectronics Reliability* 的主编，*IEEE Transactions on Components and Packaging Technology* 副主编。美国马里兰大学学院园分校高级生命周期工程研究中心（Center for Advanced Life Cycle Engineering，CALCE）创始人；机械工程系 George Dieter 首席教授、应用数学系教授。已编写关于电子产品开发、使用和供应链管理等方面的书籍 20 余本，发表学术论文 400 余篇。在过去的十年中，他领导的科研小组一直致力于可靠性诊断的研究。他们已经为 100 多家生产电子产品的跨国企业提供了咨询，为这些企业的电子产品和系统的策略制定、设计、试验、诊断分析、知识产权、风险评估等方面提供了专家意见和技术支持。因对可靠性研究的杰出贡献，获得欧洲微纳米可靠性奖；因对电子封装的研究，获 3M 研究奖；因对电子产品可靠性分析的贡献，获 IMAP（国际微电子和包装协会）William D. Ashman 成就纪念奖。

各章节作者介绍

Harold S. Balaban，在美国国防部和其他政府机构工作40余年，主要负责开发武器系统模型，并以此来实现成本和效能的分析。现受雇于美国国防分析研究所(IDA)。在此，他将可靠性和维修性概念引入到武器系统生命周期成本和效能的建模中，并开发出了一系列模型和成本估算关系，使得这些工作更为高效、精确。典型的案例有：关于维护人力资源估计的 IDA IMEASURE 项目；运输机任务能力比率模拟原型；为估计仓储水平的可补偿性和消耗成本而建立的 IDA CER 模型。在就职于 IDA 之前，他在 ARINC 研究公司工作，最后的职位是主管、高级分析师。他负责开发和应用分析模拟模型，并用分析模拟模型来研究军事系统的成本、效能、可靠性、维修性和可用性。他所领导的小组开发的“系统可测试性和维修性程序”获得了巨大成功，此程序用以改进军事系统的组织诊断技术，是同类软件的先行者。Balaban 博士还是一名把长期质量保证和后勤控制引入到军事系统采办的主要贡献者。他撰写并出版了大量关于可靠性和维修性的论文，参与过三部书籍的撰写工作，在乔治华盛顿大学、马里兰大学教授关于可靠性理论和运筹学的研究生课程。在乔治华盛顿大学获得了数理统计博士学位。他参与了本书第 1 章、第 12 章和第 15 章的撰写工作。

Ned Criscimagna，是 Criscimagna 咨询有限公司的创始人、总裁。他为工业和政府机构提供培训、项目评估和与可靠性相关的咨询服务。在创立自己的公司之前，Criscimagna 在 Alion 科学技术公司（前伊利诺伊理工学院研究所）就职，职位是高级科学顾问，此外还担任过5年的可靠性分析中心（RAC）主任等职。在 1994 年加入伊利诺伊理工学院研究所之前，他在 ARINC 研究公司任过多个职位。在创办私人公司之前，他曾在空军的工程、维修和其他部门任高级职员 20 年。在作为空军和空军司令部职员时，他帮助制定并实施了关于可靠性、维修性、质量和系统采办方面的政策方针。曾作为空军修理和维修“2000 人研究小组（English)”的成员，为国防部总体质量管理方法最初的建立做出了贡献。美国质量保证、汽车工程师协会会员，物流工程师协会高级会员。在内布拉斯加大学林肯校区获机械工程领域理学学士学位，在空军技术研究所获系统工程——可靠性工程硕士学位。他参与了本书第 1 章的撰写工作。

Diganta Das，马里兰大学学院园分校机械工程博士，印度理工学院制造科学与工程专业技术学士，高级生命周期工程研究中心研究员。他所专长的研究领域包括电子零部件可靠性、环境和运行评级；电子零部件性能提升、再加工以及相关技术发展趋势；零部件选择和管理方法等。为零部件选择和管理，以及可靠性实践活动

实施了标杆管理和电子企业组织模式。他还在设计改进方面协助过一些企业组织。在以上领域发表过50余篇学术论文，在相关国际会议和研讨会上介绍过自己的研究工作。曾作为两项IEEE标准的技术编辑，现正协助制定另外两项标准；*Microelectronics Reliability* 和 *International Journal for Performability Engineering* 的编委会委员；六西格玛黑带大师；IEEE和IMAPS会员。他参与了本书第2、4、5、9和16章的撰写工作。

Abhijit Dasgupta，马里兰大学CALCE电子封装研究所教师兼研究员。伊利诺伊大学机械理论和应用专业博士，研究领域为非均匀材料及其结构的构成和损伤行为的微观力学建模，主要研究方向为疲劳和蠕性疲劳的相互作用。其他研究内容还包括综合热力载荷下的相关应力分析，失效物理模型的设计构思，为验证测试、可靠电子封装的筛选和降格使用提供指南。他参与了本书第6章的撰写工作。

Joanne Bechta Dugan，1980年获费城拉萨尔大学数学与计算机科学文学学士，1982年、1984年分别在北卡罗来纳州达勒姆杜克大学获电子工程硕士、博士学位。Dugan博士现在是杜克大学电子工程系副教授，三角研究园区（Research Triangle Institute）访问科学家。她主要研究具备硬件和软件故障容错能力计算机系统分析技术的开发和应用。她感兴趣的领域有：硬件和软件可靠性工程，容错计算，利用动态故障树、Markov模型、Petri网和仿真技术进行数学建模等。她是IEEE的高级会员，计算机协会会员，Eta Kappa Nu（美国电气和计算机工程荣誉协会）和Phi Beta Kappa（美国大学优等生荣誉协会）会员。她参与了本书第11章的撰写工作。

Robert M. Hecht，ARINC研究公司的首席工程师。他的专业领域为特殊设备的可靠性评估、维修性和可测试性问题，产品改进项目的规划和管理。他参与了多个武器系统开发项目，包括P-3C、E-2C、bA-6E、EA-6B、ES-3、GUARDRAIL、QUICK FIX、EF-111A和M1 Abrams主战坦克。在电子产品、机电产品和纯机械系统为可靠性而设计方面具有丰富的经验。在加入ARINC之前，他是贝尔航空航天公司新奥尔良子公司的可靠性工程师。在那里，他对美国海军表面效应船和气垫船项目进行了可靠性分析，在美国军方的帮助下，他管理了普通军用设备的可靠性和维修性验证测试工作。他是宾夕法尼亚州立大学航空工程理学学士，新奥尔良大学工程理学硕士，经ASQC（美国质量管理协会）认证的可靠性工程师。他参与了本书第14章的撰写工作。

Jun Ming Hu，中国深圳微软亚洲硬件研发中心（MACH）主管，MACH小组主要负责微软硬件产品的设计、工程、检测和制造，包括鼠标、键盘、摄像头、Xbox控制器、游戏文本输入装置、Zune音乐播放器的附件和销往世界各地的其他硬件产品，还为Xbox控制器的制造、测试、零部件生产和质量检测提供支持。MACH小组管理着中国众多的设计和制造业合作伙伴。2000年，Hu博士在雷德蒙加入了微软，身份是硬件可靠性工程师主管和零部件工程师主管。2004年，他回到深圳建立了MACH机构。在加入微软前，他在密歇根的福特汽车公司工作了8年，其职位是高级技术专

家，计算机辅助汽车电子产品设计的工程主管。1982 年和 1985 年，他在上海交通大学分别获得理学学士和理学硕士学位；1989 年，在马里兰大学获博士学位。在美国和其他国家，Hu 博士注册了 14 余项关于电子产品和质量管理方法的专利。在 1993—1998 年期间，他是 *IEEE Transactions on Reliability* 副主编，*Journal of the Institute of Environmental Sciences* 1993—1998 年的编委会委员，亚裔美国人企业成就奖获得者，两项亨利·福特技术奖获得者。他参与了本书第 3 章和第 6 章的撰写工作。

Mark Kaminskiy，马里兰大学学院园分校技术和系统管理中心的首席统计员。他是可靠性统计和概率、寿命数据分析、工程系统风险分析等方面的研究者和咨询师。他的研究和咨询项目受到诸多政府机构和工业企业的资助，如美国交通部、海岸警卫队、陆军工程团、海军、核能管理委员会、美国机械工程师协会、福特汽车、高通公司以及其他工程企业。他单独撰写或作为参与者撰写了 100 余篇期刊、会刊论文、报告和书籍，主要包括 *Modeling Population Dynamics for Homeland Security Applications*；与 B. Ayyub 合著，由 J. G. Voeller 主编的 *Wiley Handbook of Science and Technology for Homeland Security*（John Wiley & Sons，2008）；与 M. Modarres、V. Krivtsov 合著了 *Reliability Engineering and Risk Analysis: A Practical Guide*（Marcel Dekker，1999，2009）；*Statistical Reliability Engineering*（John Wiley & Sons，1999）中“加速测试（参见本书第 5 章）”的相关章节；IEEE 大百科全书的 *Statistical Analysis of Reliability Data*（John Wiley & Sons，vol. 20，1999）等。他是圣彼得堡理工大学（俄罗斯）的核物理学硕士，圣彼得堡电气工业大学（俄罗斯）电机工程博士，参与了本书第 3 章的撰写工作。

Richard Kowalski，在 ARINC 工作 27 年后，他于 2002 年退休，最后的职位是产品保障部门主管。他主要负责和硬件、软件质量项目相关政策的制定和执行。受过软件能力评价培训（SEC），在几个美国和欧洲公司，还有 ARINC 中，他利用卡内基梅隆软件工程研究所开发的能力成熟模型和集成模型开展了一些软件能力评估（SCE）工作。Kowalski 是西格玛西（Sigma Xi）会员；电气与电子工程师学会（IEEE）的终身高级会员。在 IEEE 可靠性协会的管理委员会任职超过 20 年；曾是 *IEEE Transactions on Reliability* 的编辑，他在美国东北大学数学专业获理学学士学位；凯西工程学院数学专业硕士和博士学位。他参与了本书第 7 章的撰写工作。

Sony Mathew，马里兰大学学院园分校机械工程学院，高级生命周期研究中心助理研究员，正在攻读马里兰大学 A. James Clark 工程学院的机械工程博士学位。研究领域为可靠性研究、锡须（tin whisker）、电子产品的诊断和健康管理。2005 年 5 月，他在马里兰大学获机械工程专业理学硕士学位。曾在 1997 年获机械工程专业理学学士学位；1999 年获印度普纳大学 MBA 学位。他参与了本书第 8 章的撰写工作。

Carol Smidts，马里兰大学学院园分校，可靠性项目下材料和核工程学院的教授助理，1986 年，她在比利时布鲁塞尔大学获物理工程专业理学硕士学位，1991 年，在

此校同专业领域获博士学位。她的主要研究方向为：动态系统可靠性、MARKOV 分析和人类可靠性。最近正在研究软件的可靠性。她参与了本书第7章的撰写工作。

Walter Tomczykowski，ARINC 公司产品生命周期管理和运行保障部门的主管，向高级系统部的副主席汇报工作。他在马里兰大学获得了可靠性工程专业理学硕士学位；在波士顿东北大学获电子工程专业理学学士学位。过去的25年里，通过一些服务机构和许多联邦机构，如国土安全部、财政部等，他领导的专业小组参与了国防部秘书办公室、国防后勤局、国防部（DoD）在可靠性、维修性、生命周期成本、人因工程、杜绝伪造、废弃管理等方面的相关项目。作为产品生命周期管理和运行保障部的主管，Tomczykowski 主要为波士顿、安纳波利斯（包括 Patuxent 河）、马里兰、代顿、俄亥俄州、圣安东尼奥、德克萨斯州、俄克拉荷马市、巴拿马市、佛罗里达等城市的私人机构提供相关服务。他是 *DMSMS Cost Factors*、*DMSMS Program Manager's Handbook* 和 *DMSMS Acquisition Guidelines* 的主要作者。经常以主要演讲人的身份受邀参加飞机老化、DMSMS 以及其他废弃管理相关的会议，以分享他在可靠性、生命周期成本和废弃管理等方面的知识。他关于可靠性的研究工作还被 *Wiley Encyclopedia of Electrical and Electronics Engineering* 收录。参与了本书第13章的撰写工作。

Igor A. Ushakov，在莫斯科物理技术研究所任教大约15年。1989年，受邀作为乔治华盛顿大学的杰出访问教授。后来，他在乔治梅森大学和圣迭戈的加利福尼亚大学任教。他同时还在美国一些著名的公司工作，如 MCI 公司、高通公司、休斯网络系统公司、Mantech 公司。在大型远距离通信系统、数学和通信系统计算建模的可靠性与效能分析方面，他有着丰富的经验。他还是许多国际会议（举行地包括美国、俄罗斯、乌克兰、加拿大、日本、英国、法国、意大利、德国、挪威、波兰、匈牙利和保加利亚）的主席。他撰写了大约300篇关于产品运行研究、可靠性工程和理论、网络通信模型的论文，均在国际著名的数学与工程杂志上发表，Ushakov 教授用俄语、英语、德语和保加利亚语撰写了大约30本著作，其中有三本在美国出版。他的著作包括：*Histories of Scientific Insights*（Lulu，Morrisville，North Carolina，2007）、*Course on Reliability Theory*（Drofa，Moscow，2007）、*Statistical Reliability Engineering*（John Wiley & Sons，New York，1999）、*Probabilistic Reliability Engineering*（John Wiley & Sons，New York，1995）和 *Handbook of Reliability Engineering*（John Wiley & Sons，New York，1994）。他是西格玛西（Sigma Xi）、加州州立大学 Omega Rho 组织、工程荣誉学会（Tau Beta Pi）的会员，国际非正式概率和统计学者组织 Gnedenko 论坛的创始人。他参与了本书第3章的撰写工作。

David Weiss，可靠性和系统分析领域的顾问。在马里兰大学，任工程研究中心可靠性项目主管10年，与其他教师一起建立了可靠性工程方面的研究生项目。在加入马里兰大学之前，他是通用电气公司的产品可靠性主管，Booz Allen Hamilton 咨询公司的合伙人。他参与了本书第15章的撰写工作。

目　　录

第 1 章　产品效能与价值

1.1 引言

任何产品或产品系统的最终目标都是以合理的成本尽可能地实现其既定功能。通常，功能被描述为某种结果的输出，例如信息系统中令人满意的信息传输、运输系统中货物的总吨位、机载气象雷达所探测气象信息的精确性。产品满足客户需求的综合能力被称为产品效能（Product Effectiveness）。如果产品有效，那么它就能很好地实现其既定功能；反之，我们就要弥补产品的缺陷。产品价值（Product Worth）是产品实现既定功能所要求的全部成本，包括购买价格、产品运行维护费用、维修和废弃处理费用等。

1.2 影响效能的产品特征

产品效能是关于多个产品特征和外部因素的函数。对于汽车来说，可信性、安全性、易维修性和舒适性是购买者认为比较重要的因素。从汽车的价值角度来看，比较重要的因素是购买价格、运行经济性和一个好的转手价格。这些方面都比较优秀的汽车，我们一般会认为它有较高的价值。对于任何特定产品来说，想要具备较高的产品效能和价值，就要在以上这些方面有独到的表现。

优秀的产品设计和开发，要求设计团队的成员（如果合适的话，还应该包括用户）在产品生命周期中适当的阶段对所有与产品效能相关的属性进行评价和讨论。这些阶段包括：概念形成阶段、研发阶段、生产阶段、运行和废弃阶段等。对于大部分产品，特别是具有较长生命周期的产品，最高的成本在产品的运行、保障和维修方面。在整个生命周期中，早期阶段的一些工作和决策直接影响着产品的后期表现和成本。表 1.1 给出了在产品设计开发过程中，早期的决策对下游开发成本的影响。例如在产品的概念设计阶段，设计决策和其他工作只消耗了开发和生产成本的 3% ~5%，但是，此阶段的工作结果却影响着产品生命周期总费用的 40% ~60%。

表 1.2 列出了产品的一些特征，这些特征在产品的性能（Performance）、可用性（Availability）和经济可承受性（Affordability）等方面影响着产品的效能。性能是指产品的运行性、物理性和功能性，可用性是指产品处于可用状态的概率，经济可承受性指与产品开发、购买和运行相关的经济影响。

理论上来讲，权衡产品的特征可以改进产品的整体效能和价值，但这是一个极其复杂的过程。例如某汽车制造商想扩大利润，并认为实现此目标最好的方案是推出一款具

有高经济承受性、高可靠性的新车，这样就可以增加市场份额，进而实现利益最大化。经济承受性是汽车制造成本的函数，经济承受性高的汽车更易于维修，但必须以牺牲或折中制造的容易性为前提。与20年前相比，现在的汽车制造成本和维修成本有了明显下降。例如电子打火这样的新设计方法使得汽车比过去更可靠；而利用电脑进行平衡校正则解决了现在复杂的发动机和传动装置所带来的问题。好的设计团队都知道，产品的特征在某些时候相互支撑，而在另外一些时候却相互制约。因此，权衡已成为产品开发过程中不可或缺的部分。

表1.1　产品开发过程和总费用的关系

开发过程阶段	开发总经费分布	
	阶段经费消耗	影响总经费消耗的比例
概念定义	3%～5%	40%～60%
设计	5%～8%	60%～80%
试验	8%～10%	80%～90%
工艺规划	10%～15%	90%～95%
生产	15%～100%	95%～100%

表1.2　影响产品效能和价值的特征

性　　能	可　用　性	经济承受性
运行：	可靠性：	成本：
范围	无故障运行	开发或采购
速度	冗余设计或降低规格	
精度	平均失效时间	闲置或运行
易损性		
有效载荷	维修性：	维修
输出功率	易维修性（备件存取、维修耗时）	
	所需资源（人力、工具）故障检测和隔离（可测试性）	处理
物理性能：		
容量和密度		
重量	后勤保障性：	
输入功率	备用资源	
使用环境	人力培养	
	工具设备	

（续）

性　　能	可　用　性	经济承受性
功能：		
安全性	开发时间	
任务完成情况		

1.3　影响产品效能的计划因素

典型的新产品开发过程揭示了产品从最初的概念到最终生产模型的艰难历程。如果产品是一项技术创新的结果——也就是说，产品通过引入全新的功能，或者以全新的方式实现已有功能，那么这类产品的开发过程就具有里程碑的意义。市场（或者现存客户基础）决定了对新技术或是改进技术的需求。设计开发小组历经艰难完成项目任务，首要目标是保证产品能以最小的故障率实现其既定功能，且在出现故障时能迅速进行维修。这些目标都必须在可接受的开发、生产和保障预算以及日程计划内完成。

性能（Performance）、成本（Cost）和日程计划（Schedule），这三项标准给企业施加了非常大的压力。就像产品的特征需要折中才能实现其预定功能一样，计划的各项目标之间通常也需要折中。这些问题最早出现于产品开发过程中，通常在基础研究和概念验证阶段。例如由于竞争的存在，分配给开发所需技术的时间和验证概念可行性的时间通常都会被迫缩短。

对于典型的产品来说，在初步工作完成之后，需要为其建立一个产品原型。在理想状况下，产品原型要尽可能地接近于最终产品，这样做的目的是初步建立满足关键效能要求的可行性特征。模型可以是硬件或软件原型，也可以是计算机模拟系统、关键子系统或组件。原型可能看起来较为粗糙且不适合生产线制造，并且只能通过熟练的技术人员使用昂贵的设备花费相当长的时间制造出来。开发阶段早期，对制造（Manufacture）、质量（Quality）和可靠性（Reliability）的关注会节省后续开发过程中所耗费的时间和财力。随着项目的进展，改进产品可靠性会变得越来越困难，耗费的资金越来越多，时间表越来越不灵活，预算资金也会越来越紧张。虽然可靠性正逐渐受到重视，但新产品开发的第一年仍然是个非常艰辛的过程，设计师需要付出很多额外的、有些时候甚至是疯狂的工作，以确定产品失效的原因，并尝试通过对产品修改、升级或者在运行和维修过程中做出改变来消除这些不利因素。

在新产品开发（革命性变化）过程中，所涉及的重要因素也会出现在改进性设计中，或者出现于设备验证的开发项目（进化性变化）中。对革命性开发和进化性开发来说，可靠性影响着产品的效能，在项目开始时开发人员就应该考虑到这一点。

图 1.1 给出了影响产品效能的主要因素：可用性、可信性（Dependability）和能力（Capability）。可用性和可信性都以可靠性为基础，维修性（Maintainability）和后勤保障性（Logistics Supportability）则是它们的主要构成元素。

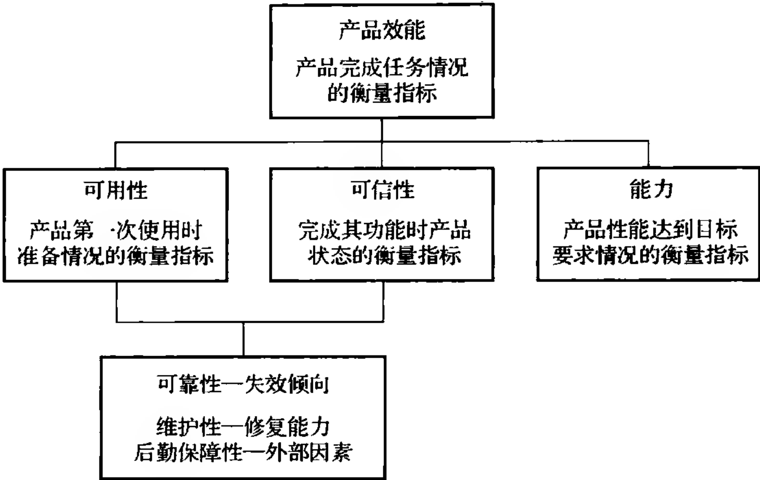


图 1.1 产品效能的主要组成部分

1.3.1 产品效能

产品效能（Product Effectiveness）可以正式定义为：在特定的状态下运行时，产品实现运行目的的能力。它受到产品的使用方式和维修方式的影响，同样也受到设计和生产过程的影响。另外，后勤保障系统、企业的政策、规章制度、产品的使用准则、财政限制以及其他行政性政策的决策也都会影响到产品的效能。

对于一次性系统（如导弹、鱼雷和熔丝）来说，运行阶段内的运行时间（Operating Time）和日历时间（Calendar Time）并不是那么重要。然而，多次使用的产品则要考虑对失效的维修，这些时间因素就十分重要，它们决定着产品的效能。如果产品在需要运行时不能工作，或者不能成功运行（即不能完成任务，或者无法实现其功能），那么产品就已失效。无论是一次性产品还是多次使用的产品，它们都要能够在用户或供应商所要求的特定环境内运行。如果产品必须在用户要求的高压环境下运行，而设计团队没有预计到这一点，那么产品的效能将会有所下降。

“特殊环境”也包括产品是连续运行还是周期性运行。当连续运行时，产品失效后才开始进行维修，任何失效都会降低产品效能。对于周期性运行的产品，如汽车或者飞机，产品不在关键期内运行时，就可以对其进行维修。有计划的预防性维修，可以消除失效隐患。每一天中抽出一些时间用以维护那些处于待命状态的产品，这样可以提高产品的效能。当然，如果在运行之前就失效的产品部分对预防性维修不敏感，那么最好使其保持连续待命状态，仅进行维护就可以了。

影响产品效能的另外一个因素是运行需求（技术性能特征）的变化。功能目标设计的改变可能会降低产品的效能。例如军事系统中如果需求额外装甲，或是额外的电子对抗，那么即便系统本身没有退化，原本用来对抗目标的系统效能也会降低。假设一辆跑车的最高设计时速为 89.4m/s，而对手的车速能够达到 93.87m/s，其他的因素都相同，那么速度较慢的车的效能就已经下降了。

有时，我们用设计效能（Design Effectiveness）和使用效能（Use Effectiveness）来描述产品的性能。设计效能用来衡量在试验条件下（如最小操作、维修和后勤影响），产品满足特定性能需求的程度。产品效能的另一个方面是使用效能，它用来评价产品达到预期要求的程度（即便这些要求超过定义范围）。虽然跑车和旅行轿车都能提供运输功能，跑车用来运载货物时，使用效能就不是很高，同样，旅行轿车的操控和加速性能也不好。

产品效能用来衡量产品满足购买者需求的程度，它是可用性、可信性和产品能力的函数。可用性是指产品能够开始工作的概率；可信性是指产品能输出满意结果的概率；产品能力是指在既定运行状态下，最初设计输出对所需任务的完成程度。

现在，我们将讨论这三个话题——可用性、可信性和产品能力，然后将讨论可用性和可信性的三个主要组成部分：可靠性、维修性和后勤保障性。

1.3.2　运行准备状态和可用性

产品在需求出现时，完成其既定功能的能力称为运行准备状态（Operational Readiness）或运行可用性（Operational Availability）[⊖]。两者的不同在于：后者仅包括运行时间和停机时间，而前者还包括产品闲置和仓储时间，也就是用户不需要产品的时间段。运行准备状态或者运行可用性与产品效能在某些方面是有所差异的，它着重于“需求出现”方面，而不是任务或使命的完成。它把焦点放在产品在某个时刻的可用性，而不是在一段时间内的可用性，如任务的完成率（成功完成任务的百分比）。产品的运行周期可能会很长，例如人造卫星在运行很长时间后才能到达另一星球；在发射的时候卫星可以运行，但不能保证在执行任务期内都维持正常运行。对于产品来说，可用性就是它在持续运行的状态下提供有用输出的能力。通常，只要计算在“需求时间”内产品能运行的概率或能提供有效输出的运行时间占总运行时间的百分比，就可以估计出可用性。

运行可用性和产品效能另外一方面的差异是后者包括设计能力，如精度、功率和重量。运行可用性却常常将与产品特征相关的细节检查排除在外，而只考虑在特定时间点产品实现其既定功能的准备程度。对于不同使用目的，一个或多个特征会影响到产品的可用性。通常，产品的运行和停机状态是客户所定义失效的函数，而用户定义的失效则取决于产品的使用状况。如果和产品关键特征相关的表现不能令人满意，用户可能就会认为产品处于停机状态，这样的话，在需求结束或者缺陷被修正之前，产品的运行准备状态或可用性就是零。

假如雷达的设计距离为 50m，而它实际只能达到 45m，我们还会认为它有效吗？如果空中避免撞击的最小距离是 50m，那么我们就可以认为安装此雷达的飞机不能飞行，此雷达不具可用性。如果 50m 是目标值，20m 是可接受的最小值，45m 的距离就可以接受。在所有能工作的时间段内，雷达的范围至少为 20m，我们可以基于这样的定义估计

⊖ 虽然，诸如可用性之类的内容曾经只与军事系统有紧密的联系，但是现在它们都已广泛应用于民用工业。例如海边的石油钻塔的可用性就十分重要。

雷达的可用性。

因此，在产品的使用环境下，运行可用性和运行准备状态与正常工作时间和停机时间有关。以下是相关定义：

① 系统或产品的运行可用性：在既定条件下，在任意时间点，产品或系统满意运行的概率。此处的总时间包括运行时间、维修实施时间、行政管理时间和保障时间。

② 系统或产品的运行准备状态：在任意时刻，产品或系统能够满意运行，或者设备在已知状态下（包括预警时间内）按用户需求运行的可能性。总日程时间是估计运行准备状态的基础。

运行可用性的子集是内在（Intrinsic）或固有（Inherent）可用性。就像设计效能的概念一样，运行可用性试图仅通过维修实施时间和用户要求的产品运行时间来最小化外部因素带来的影响。因此，它不包括自由时间。自由时间是指不需要产品运行的时间以及产品停机时间，这些都是由后勤供给和管理延迟而造成的。固有可用性是一项内在能力，因此，假设实际运行条件与设计定义的条件相匹配，设计和生产工程师首先要说明发现的问题。如果工程师不能解决问题，那么产品运行管理者就要减少行政管理、后勤延迟，或者更有效地利用并维修产品。

1.3.3　可信性

大部分产品都处于其运行过程中多种状态的某一种状态，可信性（Dependability）用来衡量产品处于某种状态的概率。如果某一产品包括 n 个可识别零部件，每个零部件都仅具有两种状态（有效和失效），那么产品就可能处于 2^n 种状态中的某一种。例如一个产品包含有10个组件，每个组件不是处于工作状态，就是处于不工作状态，那么它就具有1024种可能的状态。

通常，运行可用性能够用单一的数字量化，但可信性却不行。然而我们却可以用可信性概念来量化产品效能。尽管如此，对于简单的例子来说，我们仍可以对其可信性进行量化表示。例如对于某一简单产品，定义其1024种状态的某一子集为有效状态，如果产品的运行状态属于此子集，那么我们认为产品是可信的。当然，产品要输出可接受的结果，并不一定需要处于所有的有效状态，此时就需要考虑产品的能力，我们将在后续内容讨论它。

从分析法的角度来讲，可信性的概念描述了产品如何从一种状态转变到另一种状态。例如产品部件的失效，通常会使产品从当前状态转变到能力低下的状态。如果及时进行维修，那么产品又会回到具备生产力的状态。如果某部件的失效直接导致产品不能工作，那么在得到维修之前，产品没有任何有效输出。

1.3.4　产品能力

产品能力（Capability）用来衡量产品完成给定任务的情况。通常，它是一个依赖于状态的量。如果产品没有运行，那么它的能力便为零，不过情况不总是这样的。假如用一辆坦克来保卫领土，有可能坦克不能开火，但如果敌人看到了坦克，却并不知道它处于这种状态，那么坦克在维修期内就算完成了保卫领土的任务。另一方面，运行中的产品不一定都能像想象的那样发挥其最大能力。例如即便所有的部件都是完好无损的，

一部航拍摄像机也会因为云层太厚而拍不到清晰的画面。产品所处的每一状态都有其对应的能力量度，例如多发动机飞机的速度、攻击范围和能量消耗就取决于正在运行的发动机的数量。

用以衡量产品能力的单位取决于产品本身及其肩负的任务。对产品性能的衡量通常与输出结果有直接关系，如图片的分辨率、所能传输的信息量、所发电的千瓦数或者对敌人造成的总伤害等。当这些量比较难定义或者难量化的时候，一个有序的范围会比较有用，例如从 0 至 100，其中 100 表示最佳输出结果。在一些情况下，也可以使用概率来度量。如果把产品可能所处的每一种状态定义为有效或者是无效的，那么产品能力就是产品在有效状态运行的概率。

1.3.5　可靠性

可靠性（Reliability）是决定产品效能的一项关键特征，它是产品避免失效的能力。缺乏可靠性会最终导致产品性能降低，或者失去性能，进而影响安全，并且会引发对修复行为的需求，如诊断、修理、备件补充和维护。具有高可靠性的产品能长时间运行，那么原本用于修复的资源就可用于提升产品性能。

在与产品效能相关的表述中，满意的运行通常与一些已定义的满意输出相关联。如果产品的所有输出都在此范围内，那么产品的运行就是可靠的。但要注意，此处定义的可靠运行，不一定就能保证产品输出满意的结果。多云环境中的航拍摄像机就是个典型的例子。

观测可靠性（Observed Reliability）是在特定状态下运行的产品数量与样本总产品数量的比值。同样，可靠性可以描述为时间的函数，图 1.2 是一个可靠性函数的示例。

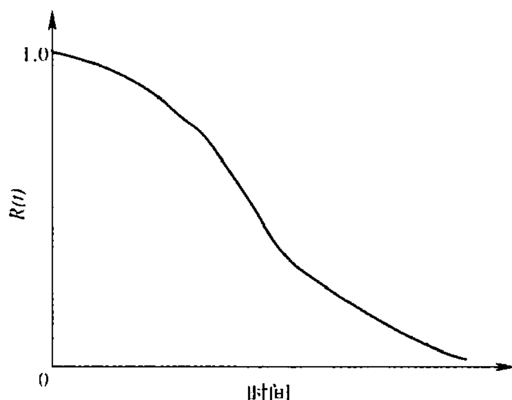


图 1.2　典型的产品可靠性曲线

在一个或多个输出超出容限（Tolerance）时，机内测试设备（Built-In Test Equipment，BITE）用来检测输出并警告操作者，我们可以持续评估拥有此类设备的产品可靠性。现在，BITE 是工业产品设计中常见的设计策略，这种策略也越来越流行于消费产品的设计，特别是电子产品。然而，在通常情况下，必须由操作者对产品的性能做出评价，区分能力问题和可靠性问题也不总是件容易的事情。有水流进了洗衣房，洗衣机的使用者通常会很容易地判断出洗衣机有问题。但是，为什么衣服没有被洗得像预想的那样干净，此问题的原因判断起来就比较复杂。无论这是可靠性的问题（例如涡轮不能适当地搅动水流），还是产品能力的问题（例如电动机能力不足以承载洗涤载荷），常见的决策是把问题归咎于可靠性，除非是对产品的效能进行了专门的分析。

假定产品在被调用时已做好了开始执行任务的准备，那么任务可靠性（Mission Reliability）一般的定义为：产品在执行任务期间成功运行的概率。因此，产品在执行任

务期间，失效的出现会导致任务不能圆满完成，任务可靠性是指不出现这类失效的概率。对于只需一次运行的产品来说，此概率是可靠性函数曲线上与任务时间相等的时间点的值；如果多次运行，产品可能会发生磨损，这时就要根据离现在最近的维修或修复调整累积运行时间或应力作用时间，在保证能完成任务的前提下，实现任务可靠性的过程要考虑所有可选的运行模式。可选模式包括使用冗余零件或备件运行，这些零部件是用来代替失效零部件的。

另一方面，无论失效何时出现，也无论它是否同时影响了任务的完成、维修需求和备件需求，后勤可靠性（Logistic Reliability）都不仅关注任务的完成，还关注所有对后勤系统有所要求的失效。在通常情况下，备件可以提高任务可靠性，但同时也可能会降低后勤可靠性。后勤可靠性的量度是需量率（Demand Rate），当一个失效触发后勤保障系统时，需量率会追踪因此而出现的需求。

1.3.6　维修性

维修行为对于把产品从失效恢复到满意状态是必不可少的，维修性（Maintainability）是指维修行为的难易程度和经济性。失效发生后，所要执行的恢复行为包括隔离失效根源、修复问题、检查产品、移除检测设备和工具、关闭所有阀门和面板、将产品恢复到可用状态以实现所要求的功能。在修复期间，产品停机时间的统计平均值称为平均故障间隔时间（Mean Down Time，MDT），它包括故障诊断时间、维修实施时间、后勤延迟和管理延迟。

使产品维持正常运行，或者失效发生时使其恢复正常的难易程度是维修性的关键特征。维修性是一个基本的设计特征，它包括所有维修实施时间。影响此特征的大部分工作都在设计阶段进行。如果要让产品非常易于维修，那么它就不能太复杂，维修设备要能很方便存取、搬运乃至替换；产品的连接件要尽可能地统一；维修所需的特殊工具要尽可能地少等，这些方面的工作由设计工程师负责。广义的维修性是指失效发生时，在特定的情况下，在给定时间内对产品进行维修，并使其恢复到正常运行状态的概率。其中，给定时间段不包括由于后勤延迟和管理延迟造成的停机时间。

可检测性（Testability）是根据失效检测和失效根源隔离定义的，它是维修性的一个子集，还包括检测和隔离的速度和精确性。理想化的情况是：只要发生了任何失效（只有失效），我们就要检测到这种情况，这样就允许操作者采取适当的操作（例如关闭产品以避免更大的损失）。失效可以通过人为观察（例如操作者可能会看到冒烟或者是无效的产品响应）来检测，也可以通过产品自身内部的检测设备来检测。同样，维修人员也可以使用手动或半自动的方法检测组件，以隔离失效根源并确认失效原因，也可以使用内置的检测设备来检测。但实际情况是：一些失效是间断性地出现，且很难被检测和隔离。

表1.3是产品所处各时间段的定义，时间是量化产品或系统特征的重要量度，因为它允许对产品的特性进行定量而非定性描述。常用的时间单位有年、月、日和小时，它们都是计算产品可靠性、维修性和可用性参数的基础单位。正是由于有这么多描述时间的单位，所以要在每次对产品进行调查时都要谨慎地选择单位，以提供所期望的结果。

表 1.3　时间元素的定义

时间元素	定　义
运行时间	产品处于操作者可接受的运行状态的时间。它包括用户对产品的运行状态不甚满意的时间，但不满意程度还不足以对产品进行停机维修。此外，它不包括维修后用户仍不满意、弃之不用的这段时间
停机时间	产品处于不可接受的运行状态，或者没有准备好运行的状态的时间之和
失效识别时间	故障情况显现之前的时间
维修实施时间	停机时间的一部分，相关人员在此时间段内进行实际的维修工作，它包括产品接受维修的准备时间和确定失效位置、修正失效、检测产品的时间
后勤延迟	停机时间的一部分，在此时间内的修理延迟（等待时间）的唯一原因是缺乏所需零部件，导致维修不可进行
管理延迟	停机时间的一部分，维修实施时间和后勤延迟之外的时间
闲置时间	用户不需要产品运行的时间段。它可能取决于停机时间、产品是否处于运行状况，也可能不取决于以上因素。在闲置期内，产品的停机时间不在操作可用性的计算范围之内
备用时间	产品可运行，但作为备用的时间段；虽然可以运行，但产品不能实现有效功能的时间段。在此时期内，产品有可能随时会被唤醒，运行
失效确认时间	从意识到失效存在到失效信息显示、试验时间和开始查找失效的时间，不包括运输或维修准备时间。访问时间反映了去掉盖子和防护物、连接试验设备的时间，很大程度上由机械设计决定
诊断时间	查找失效位置的时间，包括调整试验设备（如设定示波器或发生器参数）、实施检测（如对照维修手册检查波形）、解释信息（利用算法对检测结果分析）、证明结论和确定修复行为所耗的时间
更换时间	移除出现故障的外场可替换组件（Line Replaceable Assembly，LRA）的时间，然后以合适的方式连接、为替换件布线。LRA 是可替换的，可以不对其进行失效诊断。替换时间在很大程度上由 LRA 和机械设计特征决定，如连接件的选择
供给延迟	从确定要对零件或组件（LRA）进行维修，到这些组件到达维修技术人员手中所需的时间。供给延迟所需要考虑的因素包括维修人员用工具取出零件的时间、从供给仓库获得零件时间、从另一地点仓库接收到零件的时间和从制造商取得零件所需的时间
检查时间	确定失效已排除，产品可以运行的时间。产品有可能在检查完成前就已恢复，这时，虽然它是一个修理函数，但所有检查时间不包括停机时间；接入了新的模块后，可能需要对产品进行调整。在检查时间内，一些或者所有的校准时间（Alignment Time）都不属于停机时间窗口（Downtime Window）

通常，我们感兴趣的时间区间是产品处于工作状态的所有日历时间。在图 1.3 中，总时间被划分为可用时间（Available Time）和不可用时间（Unavailable Time）。在可用时间内，产品可以完成既定目标；在不可用时间内，产品处于供应、修理或存储状态，不可用。

因此，有两个划分时间的标准：设备的可运行状态和对它的需求状态。以下是这两个标准的图形描述。

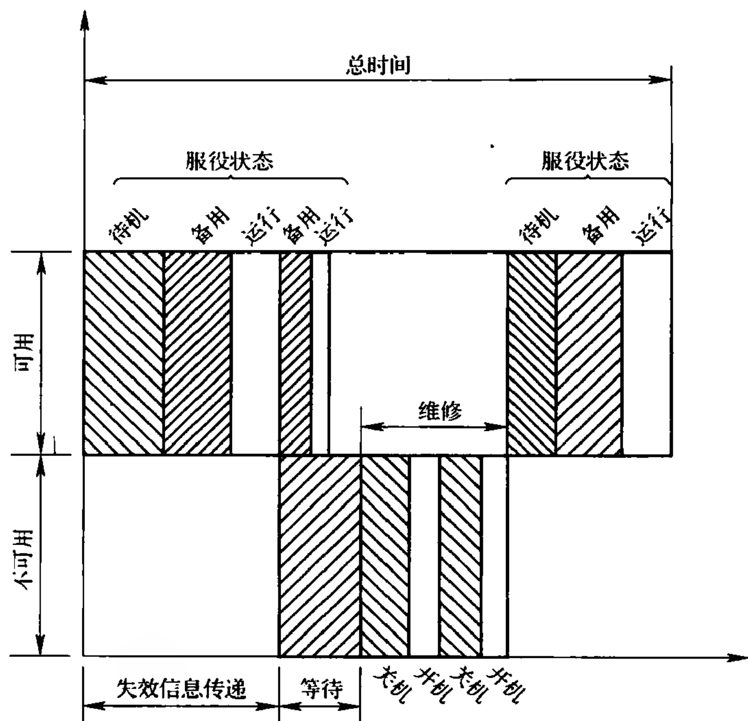


图 1.3 总时间的划分方法

标准 1：产品的可运行状态

1）产品处于可运行/不可运行状态的时间。

2）管理延迟。

3）后勤延迟。

4）失效确认时间。

5）维修时间。

标准 2：对产品运行的需求

1）需要产品运行。

2）不需要产品运行。

① 仓储时间。

② 空闲时间。

③ 备用时间。

1.3.7 时间元素之间的关系

检查各时间元素之间的关系能提供额外的产品效能组成部分的特征。在此帮助下，图 1.4 给出了各时间元素是如何组合并影响产品效能组成部分的。注意：在正常情况下，产品能力是与时间无关的参数，因此，图中没有显示时间因素对产品能力的影响。

1.4　任务目标分解

即便在讨论如何对图 1.4 中的概念进行定量衡量之前，我们也可以阐明它们如何有助于确定产品失效的根源，并分配维修任务以改进产品效能。这些概念可以提供一些信息。用于相竞争的设备或系统的比较评价，也可用于判定特定产品特征的差异所在。

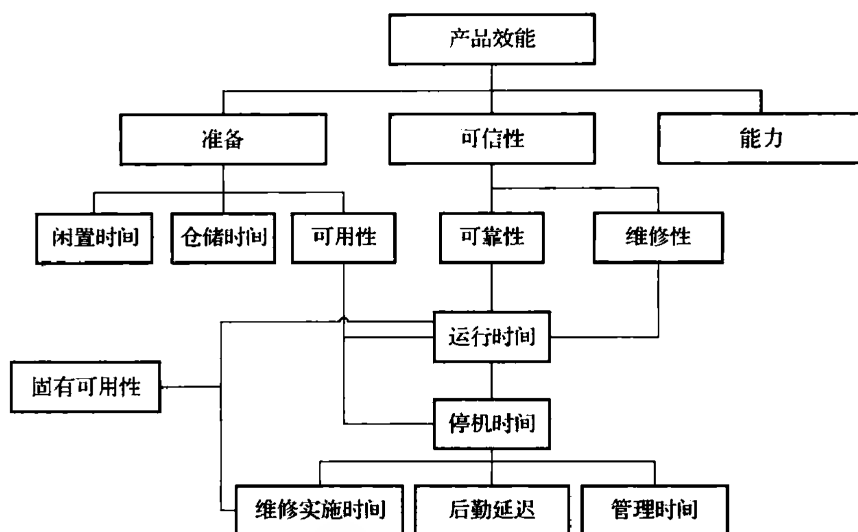


图 1.4　各时间元素之间的关系以及对产品效能的影响

时间元素是根据各时间区间的长度以及各种人员组织责任之间的关系进行划分的。显然，管理者需要对产品的空闲时间、仓储时间、管理时间和后勤延迟时间负责。生产和设计工程师要对运行时间（失效频率）、维修实施时间负大部分责任。当然，维修和设计工程师同时对维修实施时间负责。

为达到最大效能，需要最大化运行时间、最小化停机时间而不使用时间（闲置时间和仓储时间）是一个安全阀。最大化闲置时间意味着将产品使用的压力最小化；备用件导致仓储时间的产生，备用件用来在紧急情况下承载运行载荷。因为产品在仓储状态和工作状态的性能退化速率是不同的，并且一些已不能运行的设备可能会被错误地放置在仓库中，因此，在确定运行准备状态时必须考虑此时间元素。

产品的短时间运行会导致大量闲置时间，例如设备需要运行的时间相对较少，也就存在固定的需求时间表。如果工作时间是受限制而不是连续的，那么也会发生这种情况。例如银行在夜间需要锁门而白天不需要。只有操作员不在时，才需要提供自动接听服务。在没有节目转播的时候，一些电视台有规则地运行。很明显，利用闲置时间对产品进行维修，可以改进运行准备状态，因此，闲置时间能在某种程度上补偿较差的维修性和可靠性。

闲置时间和仓储时间的一个重要作用是它们提供了一种用以缓解设备不足的灵活的管理方法，并通过此管理完成了运行准备状态。但有一点非常重要，闲置时间和仓储时

间对于改进较差的设备没有任何作用。相对于获取更好的设备来说，这仅仅是无奈之举，但有些时候却是必要的选择。它们是质量的替代品，而不是达到质量的办法。

从以上的讨论可以看出，我们可以在闲置时间和仓储时间之外的时间内发现一些更重要的设备特征指标。图1.4把其他所有时间类型都与产品的可用性相关，它把运行时间和总停机时间结合了起来，还包括停机时间的三个子时间量：管理时间、后勤时间和维修实施时间，管理和工程人员对这些子时间量负责。

1.4.1 行政管理时间

行政管理时间（Administrative Time）几乎全部由管理决策和个人行为策略决定。管理决策与处理记录信息相关，个人行为策略包括维修工程师、技术人员和其他相关事务人员的个人行为策略。行政管理人员的责任就是建立有效的方法来监测、处理和分析维修行为。

除此之外，管理时间还包括被浪费掉的时间，管理人员要对此负责。它独立于工程范围之外，也与设备制造商毫无关系。

1.4.2 后勤支持时间

后勤时间（Logistic Time）是指维修中的延迟时间，此延迟是由于替换不可用零件造成的。虽然此时间在很大程度上由管理人员控制，但运行状况和设备内建的运行应力水平承受能力决定了替代零件的需求。如果得当的话，采购人员的策略能最小化后勤时间。因此，与此相关的管理者的责任与直接影响其他时间种类的管理者的责任是不同的。这也是需要对后勤时间单独考虑的原因。

1.4.3 维修实施时间和运行时间

维修实施时间（Active repair Time）和运行时间（Operating Time）都由设备内在的特征决定，因此，设备制造商主要对此负责。如果要改进维修实施时间和运行时间，要么采取行动以减少失效率，要么提高易修理性，要么同时采取这两种行动。运行时间和维修实施时间分别与可靠性和可修理性（Repairability）概念相关。在固有可用性中，这些概念自始至终都相互关联。

对于减少维修实施时间，或增加运行时间（也就是无失效时间），管理者几乎不能有所作为。但通过将运行应力级别控制在设计范围内，并保证为维修车间提供合适的工具和技术熟练人员，管理者可以影响这些时间元素。

由于广泛购买的存在，购买者总希望能以最低的总成本买到能实现预计功能的产品。关于成本的研究表明，如果在产品设计的早期能给予可靠性和维修性适当的关注，拥有产品的总成本（包括产品服务寿命内的产品启动和运行成本）会大大减少。这一发现引出了产品价值的概念，如图1.5所示。它还把产品效能和总成本、日程计划以及人力需求联系了起来。

产品价值的优化为项目管理者提出了一个难题：如何在降低总成本、开发时间和人力需求（见第13章）的情况下，最大化产品的效能。军事和商业项目的管理者都面临着成本、时间和人力等方面的限制。在商业项目中，上市时间和竞争是额外的限制。而大部分军事项目中的政治条件限制是军事项目管理者所要面对的独特问题。实际上，两

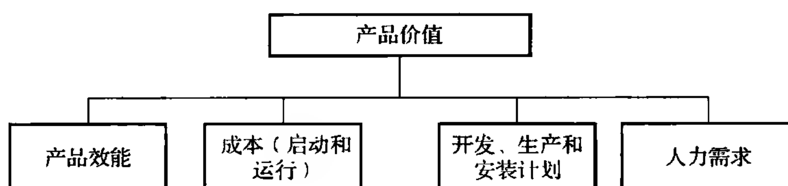


图 1.5　产品价值相关概念

种类型的管理者都会在一些可选的选项中做出决策，这些可选的选项是开发能力所能保证的产品或组件。制定技术开发计划能让此决策过程变得容易，如图 1.6 所示。此时需要说明的是，产品效能这一概念适用于在运行条件内运行的产品，它是可度量的。尽管如此，由于实际运行条件通常是不可预知的，或者超出了产品制造商的控制范围，只有某些产品效能概念的元素可以用作制定合同。从实际角度出发，必须要分析产品的任务或使用状况，以决定所需的固有可用性（Intrinsic Availability）级别，以及性能特征（设计能力，Design Capability）。如果有冗余或多模式运行存在，定义产品需求将变得更加复杂。

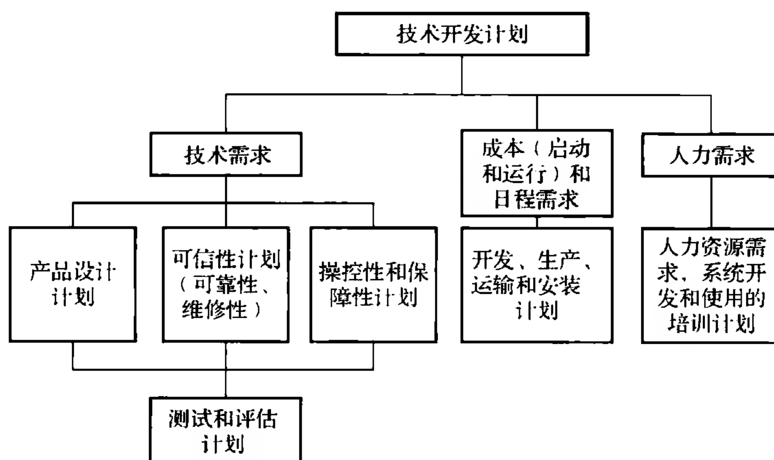


图 1.6　产品技术开发需求

例如在最佳成品水平下，为了达到需求的可用性级别，产品设计团队必须要考虑一些可选方法。是使用一些冗余系统，以便让一个或多个备用系统总是可用，失效系统也可以在最小环境下进行修理，还是开发一个具有高可靠性并能很快进行维修的产品？在很多情况下，使用高质量零部件、冗余电路和内接组件可以改进可靠性和可修理性，简单的或半自动化的失效隔离设备对于改进可用性，减少总停机时间有很大帮助。因此，在最佳财力、人力和时间的情况下，权衡分析的实质通常是决定满足用户要求的产品可用性的最小需求条件，权衡分析也需要控制时间。

电子系统中可靠性的重大历史性改进是：真空管转变为固态晶体管，再转为微芯片。设计和应力分析中新材料、新方法的使用降低了失效率。随着可靠性问题的缓解和解决，对于维护性、后勤和成本问题的关注越来越明显。

产品效能为实现产品运行、保障和性能提供了一种方法。例如在20世纪60年代早期，关注于冗余设计的可靠性工程师发现，为理论上功能完全相同的运行部件增加额外部件和一个转换机制可以改进可靠性。但这样的设计实施后，在诸如功率、重量、维修性和成本方面所带来的负面影响却更加明显。如果一个额外设计的电力供给引发了电力供给失效率上升，那么我们需要为一个重复发射电路增加什么样的值？产品效能和产品价值考虑了成本和资源的使用，如何以最佳方式指导设计和开发，以让所需的性能表现更加有效？产品效能和产品价值为此决策提供了一个概念性框架。

第 2 章　与可靠性相关的概念

2.1　引言

本章将介绍一些与可靠性相关的基本定义和数学理论。本章主要的重点是可靠性量度，其中包括：可靠度[⊖]（Reliability）和不可靠度（Unreliability）函数、概率密度函数（Probability Density Function）、故障率（Hazard Rate）、条件可靠性函数（Conditional Reliability Function）和失效时间（Time-To-Failure）等。

2.2　可靠度

设有来自于测试或监控中的大小为常数 n_o 的产品样本，在任意时刻 t，如果失效产品数量为 n_f，仍正常运行的产品数量为 n_s，则

$$n_s(t)+n_f(t)=n_o \tag{2.1}$$

公式（2.1）中的“时间”因子可以是寿命、持续总时间、运行时间、循环次数或行驶里程，也可以用一个测量值来代替，其取值范围为（$-\infty \sim +\infty$）。这个数值在统计学中称为变量（Variate）。它可以是离散的（例如循环次数），也可以是在特定区间内取的任意连续值。

产品（也可以是过程或事件）在给定时间内的失效次数是基本的可靠性指标。对于在任意时刻 t 处的产品，失效产品数与样本量的比值就是不可靠度的估计值 $\hat{Q}(t)$，也就是：

$$\hat{Q}(t)=\frac{n_f(t)}{n_o} \tag{2.2}$$

式中，变量上的“^”表示它是一个估计值。同样，把 t 时刻的产品可靠度估计值记为 $\hat{R}(t)$，它是正常运行（未失效）的产品数量与样本量的比值：

$$\hat{R}(t)=\frac{n_s(t)}{n_o}=1-\hat{Q}(t) \tag{2.3}$$

$\hat{Q}(t)$ 和 $\hat{R}(t)$ 都是一个分数，它们的取值范围为 0 ~ 1。将其乘以 100%，就可以得到相应的百分数形式。

⊖ 为了便于理解，在涉及计算和数据时，译者把“Reliability”译为“可靠度”，在其他情况下则译为“可靠性”。

案例 2.1

某半导体加工厂平均每周的产量为1000万。在过去一年中，总共有10万件产品在最终检测中不合格。

(a) 根据试验，半导体产品的不可靠度是多少？

(b) 如果检测能够剔除99%的有缺陷产品，那么，用户有多大的概率买到有缺陷的产品？

解：(a) 工厂每年的总产量为

$$n_o = 52 \times 10 \times 10^6 = 520 \times 10^6$$

同期内，被剔除的产品（有缺陷产品）数量为

$$n_f = 1 \times 10^5$$

因此，根据公式（2.3）可以计算产品不可靠度为

$$\hat{Q}(t) = \frac{n_f(t)}{n_o} = \frac{1 \times 10^5}{520 \times 10^6} \approx 1.92 \times 10^{-4}$$

即产品有1/5200的概率失效。

(b) 如果被剔除的产品是有缺陷产品的99%，那么通过测试的有缺陷的产品数量为

$$x_d = \left[\frac{1 \times 10^5}{0.99} - (1 \times 10^5)\right] \approx 1010$$

因此，客户买到有缺陷产品的概率，或者售出产品在初次使用时的不可靠度为

$$\hat{Q}(t) = \frac{1010}{(520 \times 10^6) - (1 \times 10^5)} \approx 1.94 \times 10^{-6}$$

即有1/515000的概率会出现失效。

通过样本测试或监测得到的产品可靠性估计值往往都具有变异性。例如设计寿命为10000h的灯泡被同时安装在同一房间里，这些灯泡不可能在同一时间，或者说都恰好在10000h的时候失效。与产品的期望寿命值一样，产品的可靠性也存在一定的差异。实际上，产品的可靠性估计通常是对这些差异的估计和度量。如果时间相同，增加样本量 n_o 可以提高可靠性估计的精确度。与抛硬币和掷骰子一样，为了通过试验测量事件发生的概率，需要有大量的样本。也就是说，要计算公式（2.2）和（2.3）中 $\hat{R}(t)$ 和 $\hat{Q}(t)$ 的真实值，就需要有无穷大的样本量。因此，可靠度和不可靠度的实际意义就是：在大量的重复性试验中，使未失效和失效的发生概率分别近似于它们的估计值 $\hat{R}(t)$ 和 $\hat{Q}(t)$。

对特定产品的参数进行一系列度量，其结果可以表示为一个柱状图，用来评估可靠性的变化。例如表2.1列出了将251个样本分成11组进行测试所得的一系列失效时间数据。把这些数据按运行时间排序放在表2.2的前两列，就可以创建如图2.1所示的柱状图。图2.1中的每一个矩形柱表示特定时间内的失效次数。此柱状图描绘了产品的寿命特征曲线（Life Characteristic Curve）。

表 2.1 对 251 个样本进行测试得到的失效时间数据

数据										
1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1
1	1	2	2	2	2	2	2	2	2	2
2	3	2	2	3	3	3	3	3	3	3
3	3	3	3	3	3	3	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6
6	6	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	9	9	9
9	9	9	9	10	10	11	11	11	11	11
11	12	12	12	12	12	12	13	13	13	13
13	14	14	14	14	15	15	15	15	15	15
16	16	16	16	17	17	17	17	17	18	18
18	18	18	18	18	18	19	19	19	19	20
20	20	20	21	21	22	22	23	23	24	24
25	25	26	26	27	27	27	28	28	28	28
28	28	29	29	29	29	29	29	30	31	31
32	32	33	33	34	34	35	35	36	36	36
36	37	38	39	41	41	42	42	43	44	45
46	47	48	49	49	51	52	53	54	55	56
58	59	62	64	65	66	67	69	72	76	78
79	83	85	89	93	97	99	105	107	111	115
117	120	125	126	131	131	137	140	142	—	—

表 2.2 对表 2.1 中数据进行分析计算后得到的数据

运行时间 /h	失效次数 (Δn_f)	存活产品数⊖(n_s)	概率密度函数 $f(t)$	可靠度 (R) ($n_s=251$)	平均故障率估计值 ($\Delta t=10$)
0 ~ 10	105	146	0.418	0.58	0.04
11 ~ 20	52	94	0.207	0.372	0.03
21 ~ 30	28	66	0.112	0.26	0.03
31 ~ 40	17	49	0.068	0.192	0.02
41 ~ 50	12	37	0.048	0.144	0.02
51 ~ 60	8	29	0.032	0.112	0.02
61 ~ 70	6	23	0.024	0.088	0.02
71 ~ 80	4	19	0.016	0.072	0.01
81 ~ 90	3	16	0.012	0.06	0.01
91 ~ 100	3	14	0.012	0.052	0.02
101 ~ 110	2	12	0.008	0.044	0.01
>110	10	0	0.043	0	—

⊖ 译者注：存活产品是指那些在特定时间内未检测出失效的产品。

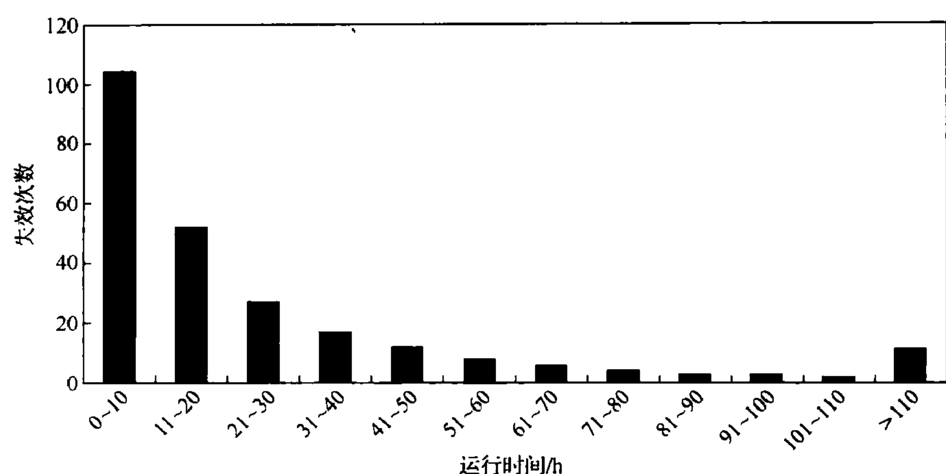


图2.1 由表2.2中数据得到的频率直方图

表2.2第5列是存活产品数量与产品总数量之比（即每个时间段结束后的可靠度），所得结果如图2.2所示。随着样本量的增加，柱状图时间间隔逐渐缩短，曲线就趋于光滑。如果失效时间是连续的，用坐标值来代替矩形柱，就可以生成光滑的故障率曲线。

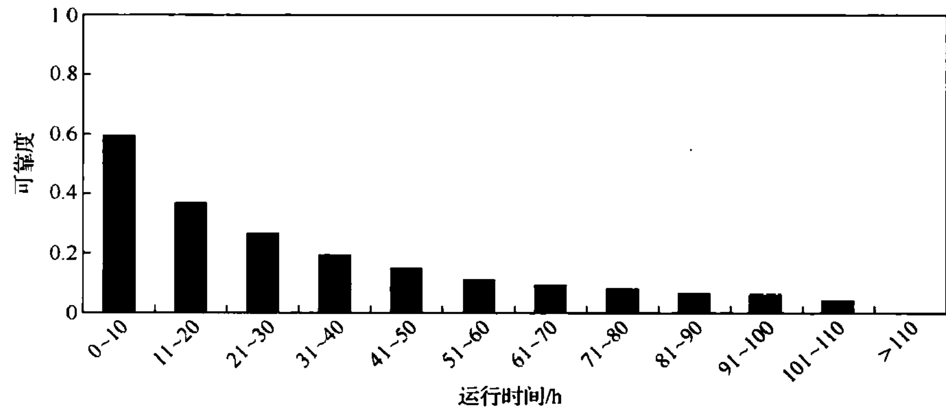


图2.2 由表2.2中数据得到的可靠度

2.3 概率密度函数

某时间段内，概率密度函数（Probability Density Function）可以根据产品失效数量与样本总量之比来估计。对于表2.1中的数据，每个时间段内的概率密度函数的估计值就是表2.2第4列的数据。图2.3的概率密度函数柱状图就是由表2.1中的数据得到的，这些概率值之和为1（也就是表2.2第4列的数据）。

概率密度函数可以表达为

$$f(t)=\frac{1}{n_o}\frac{\mathrm{d}[n_f(t)]}{\mathrm{d}t}=\frac{\mathrm{d}[Q(t)]}{\mathrm{d}t} \tag{2.4}$$

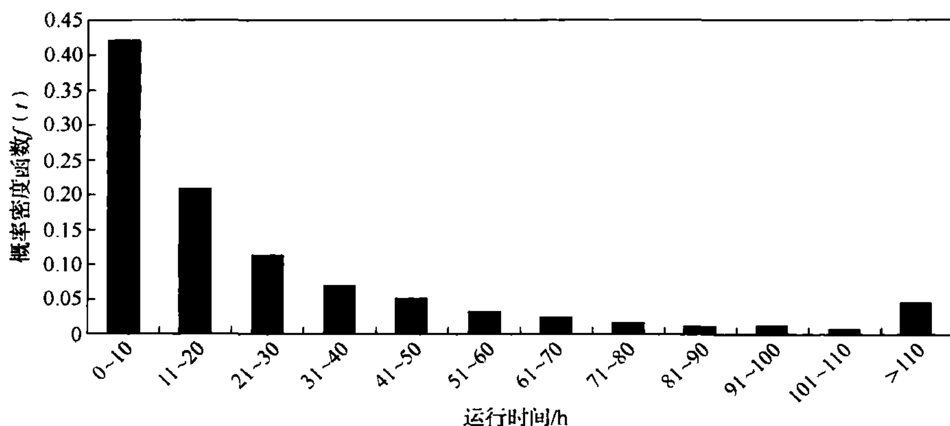


图 2.3　由表 2.1 数据得到的概率密度函数

对等式两边进行积分，就可以得到不可靠度关于函数 $f(t)$ 的表达式

$$Q(t) = \frac{n_f(t)}{n_o} = \int_0^t f(\tau)\mathrm{d}\tau \tag{2.5}$$

其中，该积分表示产品在 τ（$0<\tau<t$）时间段内发生失效的概率，即概率密度函数曲线下 t 时刻的左方区域内的面积。我们也可以把不可靠度称为连续随机变量的累积失效概率分布函数。

同样，函数曲线中 t 时刻右下方区域面积表示到 t 时刻没有失效的产品概率，可以由下式表示：

$$R(t) = \int_t^\infty f(\tau)\mathrm{d}\tau \tag{2.6}$$

因为对于一个样本总体来说，在产品寿命终结时，其失效概率总和为 1，所以，函数 $f(t)$ 也可以记为

$$\int_t^\infty f(t)\mathrm{d}t = 1 \tag{2.7}$$

案例　2.2

由柱状图 2.3 计算：

（a）产品在第 30h 处的不可靠度。

（b）产品的可靠度。

解：由柱状图可知，对于离散型数据，其不可靠度是失效概率密度函数从 $t=0$ 到 $t=30$的所有值之和为 74%，可靠度就是密度分布函数从 $t=30$ 到 $t=\infty$的所有值之和为 26%。可靠度和不可靠度之和总是等于 1。

2.4　故障率

对于不可修复产品来说，故障率（Hazard Rate）是产品在给定时间段内（从 t_i 到

$t_i+\Delta t$）失效的条件概率。假如产品存活到 t 时刻，那么其故障率为

$$h(t)=P(t_i\leqslant\tau\leqslant t_i+\Delta t\tau\geqslant t_i) \tag{2.8}$$

假定产品存活到 t_i 时刻，Δt 时间内平均故障率 $\hat{h}$ 可以表示为

$$\hat{h}=\frac{1}{n_{bp}(t_i)}\frac{\Delta n_f}{\Delta t} \tag{2.9}$$

其中，$n_{bp}(t_i)$ 表示在此时间段开始处被监测或测试的有效产品；Δn_f 为抽样阶段内的失效次数；Δt 为抽样时间间隔。随着 n_{bp} 的增大，抽样时间间隔将接近于0，故障率的平均值将接近（或就是）产品在时间 t 处的瞬时故障率。也就是说，Δt 趋于极限值0时，公式（2.9）就变成

$$h(t)=\frac{1}{n_s}\frac{dn_f(t)}{dt} \tag{2.10}$$

故障率 $h(t)$ 是每个未失效产品在单位时间内的失效概率。因为它与样本空间无关，所以它是相对失效率。基于公式（2.2）、公式（2.3）和公式（2.10），故障率和可靠度之间的关系为

$$h(t)=\frac{-1}{R(t)}\frac{dR(t)}{dt} \tag{2.11}$$

对公式（2.11）在运行时间 $0\sim t$ 范围内进行积分，并对等式两边进行指数运算［注意：$R(t=0)=1$］：

$$R(t)=\exp\left(-\int_0^t h(\tau)d\tau\right) \tag{2.12}$$

这是用故障率表示可靠度的基本公式。结合公式（2.4）、公式（2.11）和公式（2.3），故障率也可以表示为失效概率密度函数与可靠度之比：

$$h(t)=\frac{f(t)}{R(t)} \tag{2.13}$$

利用表2.1中的数据和公式（2.9）可以估算出（Δt 时间段内的）故障率，结果在表2.2最后一列给出。图2.4是故障率相对于时间的柱状图。

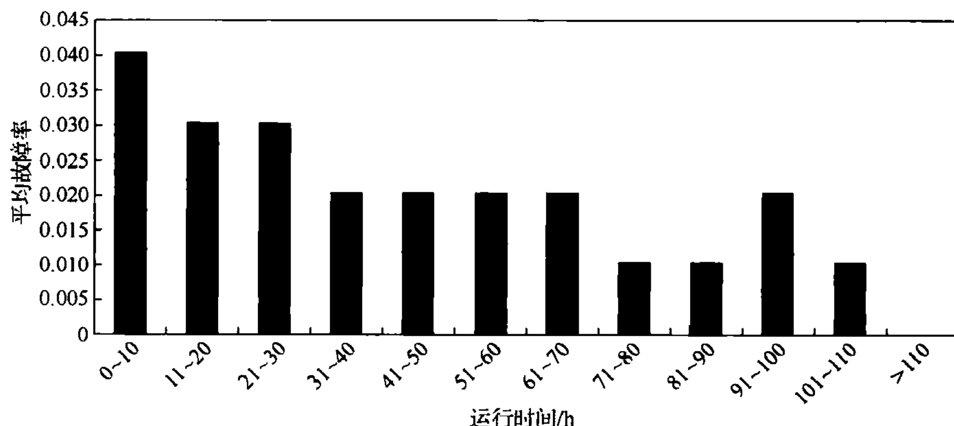


图2.4　由表2.1数据得到的故障率柱状图

2.5　条件可靠性

条件可靠性函数（Conditional Reliability Function）$R(t,\ T)$ 的定义为：假定不可修复产品已经运行的时间长度为 T，产品再正常运行一段时间 t 的概率。条件可靠性函数可表述为在（$t+T$）时刻的可靠度与 T 时刻的可靠度之比：

$$R(t,T)=\frac{R(t+T)}{R(T)} \tag{2.14}$$

其中，T 为新测试或任务开始的时间，即产品的"年龄"。

对于故障随 T 降低的产品，条件可靠性会随"年龄" T 的增加而升高；而对于故障随 T 增加的产品，其条件可靠性将降低。对于恒定失效率的产品，条件可靠性独立于 T；也就是说，任务时间 t 内的可靠度和先前的运行时间无关。这就表明，无论在任何时候，我们都可以把具有恒定失效率的产品当做新产品看待。

案例　2.3

假定某系统的可靠度函数服从指数分布，即

$$R(t)=e^{-\lambda_0 t}$$

其中，λ_0 是常数（即恒定故障率）。假定系统已经服役 10 年，计算系统在任务时间 t 内的可靠度。

解：利用公式（2.14）

$$R(t,10)=\frac{R(t+10)}{R(10)}=\frac{e^{-\lambda_0(t+10)}}{e^{-\lambda_0 10}}=e^{-\lambda_0 t}=R(t)$$

结果表明，系统的可靠性和新产品一样，而与其"寿命"无关。

2.6　失效时间

中值（Median，M）是概率分布的中间值，它将分布区域从中间（即可靠度为 50% 的时刻）划分为两个相等的部分，即

$$\int_0^M f(t)\,dt = 0.5 \tag{2.15}$$

M 是一个界限，我们很难决定它的显式关系。所以平均值是一个更好的量度。

平均失效时间（Mean Time To Failure，MTTF）被定义为失效概率密度函数的期望值，其表达式为

$$\mathrm{MTTF} = \int_0^\infty tf(t)\,dt \tag{2.16}$$

公式（2.16）等价于

$$\mathrm{MTTF} = \int_0^\infty R(t)\,dt \tag{2.17}$$

只有在失效分布函数已知时，才能使用 MTTF，因为用 MTTF 表达的可靠性函数是由失效数据建立的概率分布函数计算而来的。实际上，拥有不同可靠度和失效分布的产品，却可能有相同的 MTTF 值。

产品或系统的早期失效，往往对安全性、质保和保障性都有着较大的影响，同时也影响着产品可带来的利润。因此，失效分布的起始阶段是可靠性关注的重点。我们通常需要估计产品前 1% 或 5% 失效的出现时间。

练习

2.1　参考表 2.1 的格式，记录 30 个回形针来回弯折 90°直致失效的数据，并计算它们的可靠度，绘制其寿命特征曲线，并分别绘制概率密度函数、可靠度柱、不可靠度和故障率的柱状图。你是否认为结果取决于回形针弯曲的次数，解释你的观点。

2.2　证明：$\mathrm{MTTF}=\int_0^\infty tf(t)\,\mathrm{d}t=\int_0^\infty R(t)\,\mathrm{d}t$

2.3　某失效概率密度函数为 $f=\dfrac{\mu^k e-\mu}{x!}$（$x=0,1,2,3,\cdots$），计算相应的 MTTF。

2.4　假定失效概率密度函数如下：

$$f(t)=\begin{cases}0 & t<t_1\\ \dfrac{1}{t_2-t_1} & t_1\leqslant t\leqslant t_2\\ 0 & t>t_2\end{cases}$$

计算相应的 MTTF。

2.5　假定案例 2.3 中的系统为一辆轿车，案例的结果是否有意义？为什么？举出一些更加适用案例 2.3 结果的产品。

2.6　假定：

$$\hat{f}=\frac{1}{n_o}\frac{\Delta n_f}{\Delta t}$$

$$f(t)=\frac{1}{n_o}\frac{\mathrm{d}[n_f(t)]}{\mathrm{d}t}=\frac{\mathrm{d}[Q(t)]}{\mathrm{d}t}$$

$$Q(t)=\int_0^t f(\tau)\,\mathrm{d}\tau,R(t)=\int_0^\infty f(\tau)\,\mathrm{d}\tau$$

证明：$\int_0^\infty f(t)\,\mathrm{d}t=1$。

2.7　假定：

$$\hat{h}=\frac{1}{n_{bp}(t)}\frac{\Delta n_f}{\Delta t}$$

$$h(t)=\frac{1}{n_s(t)}\frac{\mathrm{d}[n_f(t)]}{\mathrm{d}t}=\frac{-1}{R(t)}\frac{\mathrm{d}[R(t)]}{\mathrm{d}t}$$

$$R(t) \equiv e^{-\int_0^t h(t)\mathrm{d}t}, h(t) \equiv \frac{f(t)}{R(t)}$$

证明故障率公式。

2.8　根据条件可靠性讨论一台计算机的失效数据（例如假定一台计算机已经在90℃下工作了 12 个循环，在此温度下，它还能继续工作 5 个循环的概率）。

$$R(t,T) = \frac{R(t+T)}{R(T)}$$

其中，$R(t, T)$ 是条件可靠性概率，即产品已经工作了 T 时间，它还能在继续工作 t 时间的概率。

2.9　如果故障率为一常数，如何简化条件可靠性公式?

提示：$R(t) \equiv e^{-\int_0^t h(t)\mathrm{d}t}$。

2.10　证明平均失效时间（MTTF）的公式：

$$\mathrm{MTTF} = \int_0^\infty tf(t)\mathrm{d}t = \int_0^\infty R(t)\mathrm{d}t$$

$$\int_0^\infty tf(t)\mathrm{d}t = \int_0^\infty R(t)\mathrm{d}t$$

第3章　统计推论概念

3.1　引言

只有从具体试验或实际使用中获得了实际统计数据，才能以此为基础用概率的形式估计产品的实际可靠性。恰当地选择分布函数的形式或类型需要用到这些统计数据，而且它们也提供了一些信息，用以确认可靠性分析中所使用模型的不同概率参数相关的先验假设。统计推论（Statistical Inference）包含了用于分析观测数据概率模型的技术。

3.2　统计估计

参数估计分为点估计（Point Estimation）和区间估计（Interval Estimation）两种：点估计提供来自一组观测数据的单个值，以代表潜在分布的一个参数或其他特征。点估计不能给出与估计精度相关的任何信息。区间估计（Interval Estimation）可构造一个置信区间（Confidence Interval），此区间含有指定置信水平的参数的真实值。

3.2.1　点估计

参数的估计必然基于一系列样本值 X_1，…，X_n 进行。假如这些值都是离散的，且它们的潜在分布不会因样本值的改变而改变，那么它们产生一个大小为 n 的随机样本，此样本来自于对随机变量 X 分布的观测。设总体分布中有未知参数 θ，我们认为随机变量 $t(X_1,\cdots,X_n)$ 是样本 $\{X_1,\ \cdots,\ X_n\}$ 的一个单值函数，也可认为它是一个统计量（Statistic）。从样本中选择一个合适的统计量并计算出它的值，就可以获得一个点估计，所选择的统计量称为估计量（Estimator）。

对于任意 θ 值，如果 $E(t(X_1,\cdots,X_n))=\theta$，则称估计量 $t(X_1,\cdots,X_n)$ 为 θ 的无偏估计量（Unbiased Estimator）。偏差（Bias）是估计的期望值和参数本身值之间的差异——偏差越小，估计量越好。

另一个需要用到的估计量 $t(X_1,\cdots,X_n)$ 的性质是一致性（Consistency）。如果估计量 t 是一致的，则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P(|t-\theta|<\varepsilon)=1 \tag{3.1}$$

它意味着：随着样本大小 n 的不断增大，. 估计量 $t(X_1,\cdots,X_n)$ 越来越接近于被估计参数 θ 的真实值。

在一些情况下，可能会有多个无偏估计量。其中，具有最小方差的估计量是最好的估计量。在参数 θ 的所有无偏估计量中，把具有最小方差的无偏估计量 t 称为有效（Efficient）估计量。

另一个有用的估计量的性质是充分性（Sufficiency）。如果参数 θ 的一个统计量 $t(X_1, \cdots, X_n)$ 包含样本 $\{X_1, \cdots, X_n\}$ 中 θ 的所有信息，则称这个统计量为参数 θ 的充分统计量。

下面主要介绍一些常见的点估计方法。

1. 矩估计法

矩估计法（Moment Method）是在随机变量的经验性矩估计的基础上估计分布函数（Distribution Function）的未知参数的方法。矩估计量等同于相应的分布矩，方程的解就是分布参数的估计量。

因为均值和方差是 X 和 $(X-\mu)^2$ 的期望值，那么一个大小为 n 的样本 $\{X_1, \cdots, X_n\}$ 的预期均值和样本方差可以分别用下面公式定义：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.2}$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{3.3}$$

$\overline{X}$ 和 S^2 分别为分布均值 μ 和方差 σ^2 的点估计。公式（3.3）中的方差估计量是有偏差的，但是通过乘以 $n/(n-1)$ 可以去掉此偏差：

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{3.4}$$

这就是方差的无偏估计量。通过对比公式（3.3）和公式（3.4）可以发现：当样本空间很大的时候，这两个估计量的差别很小。

案例 3.1 某装置的寿命 T 是一个服从指数分布的随机变量：

$$f(t) = \lambda e^{-\lambda t} \tag{3.5}$$

通过加速寿命试验，测得的失效时间分别为 22h、24h、31h、41h、52h、63h 和 70h。为了求解分布的参数 λ，把这组数据当作一个大小为 7 的样本 t。由于指数分布是单参数分布，所以它的一阶矩为

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{1}{7}\sum_{i=1}^{7} t_i = 43.3\text{h} \tag{3.6}$$

这样就得到均值的一个估计量。均值和参数 λ 之间的关系为

$$\theta = \int_0^{\infty} e^{-\lambda t}dt = \frac{1}{\lambda} \tag{3.7}$$

因此，参数 λ 的一个估计量为 $\bar{\lambda} = 1/\bar{t} = 0.0231$（1/h）

2. 极大似然估计法

极大似然估计法（Method of Maximum Likelihood）是一种最常用的估计方法。设随机变量 X 的密度函数为 $f(x, \theta_0)$，其中，θ_0 为参数。用极大似然估计法可以得到 θ_0 值，θ_0 有最大（或最可能）的概率（或概率密度）产生特定大小的集合 $\{X_1, \cdots, X_n\}$。获得特定样本集合的似然性与在样本点 $X_1, \cdots, X_n$ 处所计算的密度函数 $f(x, \theta_0)$ 成正

比。通常，把似然函数记作

$$L_f(X_1,\cdots,X_n;\theta)=f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta) \tag{3.8}$$

极大似然函数的基础是 n 个事件 $X=X_1$，…，$X=X_n$ 联合发生的概率（离散型问题）或密度（连续型问题）。极大似然估计 $\hat{\theta}_0$ 是使得似然函数 $L_f(X_1,\cdots,X_n;\theta_0)$ 获得最大值的 θ_0 的值。

应用一般的求极值法则，将似然函数对 θ_0 求偏导并使式子等于0，得到：

$$\frac{\partial L_f(X_1,\cdots,X_n;\theta_0)}{\partial\theta_0}=0 \tag{3.9}$$

若能验证方程的一组关于 θ_0 的解是似然函数 $L_f(X_1,\cdots,X_n;\theta_0)$ 的最大值点，那么这组解就是极大似然估计 $\hat{\theta}_0$。

由于似然函数存在倍增性，通常用求似然函数对数的最大值来代替，这样比较方便，即

$$\frac{\partial\log L_f(X_1,\cdots,X_n;\theta_0)}{\partial\theta_0}=0 \tag{3.10}$$

这个公式的解 $\hat{\theta}$ 和公式（3.9）得到的解是相同的。对于分布密度函数有两个或者更多参数的情况，似然函数记作：

$$L_f(X_1,\cdots,X_n;\theta_1,\cdots,\theta_m)=\sum_{i=1}^{n}f(X_i,\theta_1,\cdots,\theta_m) \tag{3.11}$$

其中，θ_1，…，θ_m 是所要估计的 m 个参数。在这种情况下，极大似然估计量可以通过求解下面 m 个方程得到：

$$\frac{\partial L_1(X_1,\cdots,X_n;\theta_1,\cdots,\theta_m)}{\partial\theta_j}=0,j=1,\cdots,m \tag{3.12}$$

通常情况下，极大似然估计是一致的，越渐近于正态分布，这种方法越有效。

让我们来估计二项分布的参数 p，在此情况下：

$$L_f(m|n)=\binom{n}{m}p^m(1-p)^{n-m},\quad m=0,1,\cdots,n \tag{3.13}$$

那么：

$$\frac{\partial\text{Log}L_f}{\partial p}=\frac{n}{p(1-p)}\left(\frac{m}{n}-p\right) \tag{3.14}$$

可知参数 p 的极大似然估计为 $p=m/n$。

同样，我们可以很容易地证明正态分布的极大似然估量均值就是样本的均值。

案例 3.2 对于案例3.1中设备的寿命试验数据，用极大似然估计法估计分布的参数。

此问题的极大似然函数为

$$L_f(t_1,\cdots,t_7,\lambda)=\sum_{i=1}^{7}f(t_i,\lambda)=\lambda^7e^{-\lambda\sum_{i=1}^{7}t_i} \tag{3.15}$$

求导后，可得：

$$\frac{dL_f(t_1,\cdots,t_7,\lambda)}{d\lambda}=[7\lambda^6-\lambda^7\sum_{i=1}^{7}t_i]=0 \tag{3.16}$$

解方程后，得到：

$$\bar{\lambda} = \frac{7}{\sum_{i=1}^{7} t_i} = 0.0231 \tag{3.17}$$

$\bar{\lambda}$ 即为参数 λ 的估计量。在此案例中，矩估计和极大似然估计得到的结果是一致的。

另外一种常用的估计方法是将在第 3.4.1 节中介绍的最小二乘法（Least Squares Method）。

3.2.2 区间估计

设 $L(X_1, \cdots, X_n)$ 和 $U(X_1, \cdots, X_n)$ 为两个统计量，它们使得参数 θ_0 在一个区间的概率为

$$P\{L(X_1,\cdots,X_n) < \theta_0 < U(X_1,\cdots,X_n)\} = 1-\alpha \tag{3.18}$$

我们就称随机区间 $[L, U]$ 为参数 θ_0 的一个 $100(1-\alpha)\%$ 置信区间（Confidence Interval），端点 L 和 U 为 θ_0 的置信界限（Confidence Limit），$1-\alpha$ 为 θ_0 的置信系数（Confidece Coefficient）。α 的常用值为 0.1、0.05 和 0.01。

如果 $\theta_0 > L(\theta_0 < U)$ 的概率等于 1，那么称 $U(L)$ 是 θ_0 的单边置信上（下）限或者单边置信界限。一个 $100(1-\alpha)\%$ 置信水平的未知参数 θ_0 是：如果对服从相同分布的随机样本进行一系列重复试验，计算每个样本的置信区间，所得置信区间的 $100(1-\alpha)\%$ 将包括 θ_0 的真值。

下面的示例介绍了构成置信界限的基本原则（另外一些广泛应用于可靠性数据分析的区间估计的应用将在第 3.4.4 节中介绍）。

下面介绍为一个已知方差正态分布的均值构造置信区间的步骤。设 $X_1, X_2, \cdots, X_n$ 是服从正态分布 $N(\mu, \sigma^2)$ 的随机样本，其中，μ 是未知参数，σ^2 已知。我们很容易得知：样本均值服从正态分布 $N(\mu, \sigma^2/n)$，因此，$(X-\mu)\sqrt{n}/\sigma$ 服从标准正态分布。这意味着：

$$P\left(-z_{1-(\alpha/2)} \leqslant \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leqslant z_{1-(\alpha/2)}\right) = 1-\alpha \tag{3.19}$$

其中，$z_{1-(\alpha/2)}$ 是标准正态分布 $N(0, 1)$ 的第 $100(1-\alpha/2)$ 个百分位数。求解括号中的不等式，公式（3.19）可以写为

$$P\left(\bar{x} - z_{1-(\alpha/2)}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + z_{1-(\alpha/2)}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \tag{3.20}$$

所以，对于已知 σ^2 的正态分布，均值 μ 的对称 $(1-\alpha)$ 置信区间是

$$\left[\bar{x} - z_{1-(\alpha/2)}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-(\alpha/2)}\frac{\sigma}{\sqrt{n}}\right] \tag{3.21}$$

置信水平 $(1-\alpha)$ 越高，置信区间越大。对于相同的置信水平 $(1-\alpha)$，其置信区间会随着 σ 的下降或者 n 的增加而变小。

3.3 假设检验

研究人员通常需要根据可观测数据来确定概率分布。当视觉化地对比几个假设密度

函数时，数据的直方图可以为如何选择分布模型提供思路。

某些统计检验（如拟合优度检验）可以否定或接受根据经验确定的假设概率分布和根据先前假设经过理论推导得出的概率假设。当两个或更多分布可用时，统计检验就可以确定不同分布模型的相对有效性。本节描述如何运用频率直方图（Frequency Histogram）研究分布模型和如何用拟合优度检验（Goodness-Of-Fit Test）来验证结果。我们将对两种常用拟合优度检验——卡方（χ^2）检验和 Kolmogorov-Smirnov（K-S）检验进行详细讨论。

3.3.1 频率直方图

频率直方图是一个随机变量的图形化、经验性描述。为某一特定的测试数据建立直方图的步骤如下：

① 从观测试验数据中选择一个足以包含最大值和最小值的范围。

② 将该范围划分成长度相等的一致性区间 Δx（根据所关注领域和数据范围的不同，划分方法会有所不同）。

③ 计算每个区间内的测量次数并画出矩形条，其高度表示区间内观测值次数（有时测量次数必须和 Δx 的长度相关）。

另外，矩形条的高度也可以由区间内的部分观测次数（相对频率）与区间长度的比值确定，即

$$f_n = \frac{\left(\dfrac{N_{x,x+\Delta x}}{n}\right)}{\Delta x} \tag{3.22}$$

式中，$N_{x,x+\Delta x}$是区间（x，$x+\Delta x$）内的测量次数，n 是总测量次数（样本大小）。

频率直方图可用来对比经验频率分布与理论密度函数的差异。如果一个假设分布的理论密度函数与频率直方图具有相同的形状（大致相同），并且理论密度函数曲线与频率直方图中垂直条的顶端非常接近，那么这个假设分布就可以用来为实际情况建模。图3.1显示了如何用频率直方图和概率密度分析某种单片微波集成电路（Monolithic Microwave Integrated Circuit，MMIC）装置的寿命（假定 MMIC 装置的寿命分布为正态分布）。

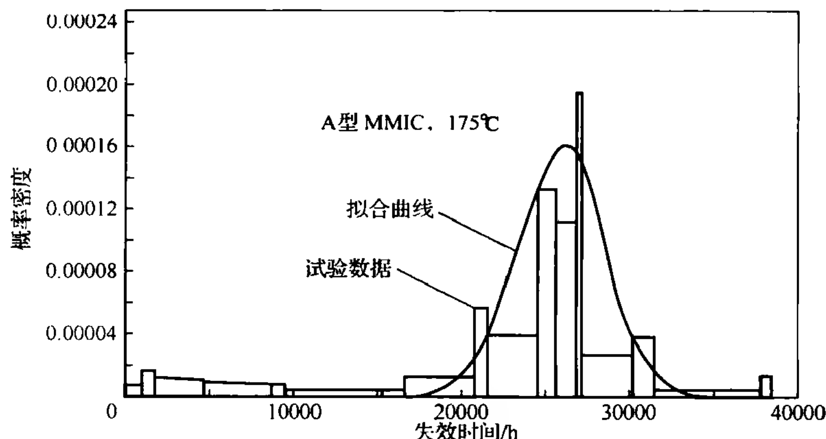


图3.1 单片微波集成电路（MMIC）装置寿命分布的频率直方图和概率密度

3.3.2　拟合优度检验

当基于一个随机变量频率直方图的大体形状为其用假设理论分布进行建模时，关于数据对模型拟合的程度有多好，没有定量评价。拟合优度检验提供了一个否定（或肯定）假设分布的定量技术。接下来将讨论两种常用检验——卡方（χ^2）检验和 Kolmogorov-Smirnov（K-S）检验。

1. 卡方检验

假定某样本的随机变量有 N 个观察值（测量值）。卡方拟合优度检验（Chi-Square Goodness-Of-Fit Test）用来比较随机变量的 k 个区间内的观测频率 n_1，n_2，…，n_k 与假设理论分布 $F_0(x)$ 中相应的频率 e_1，e_2，…，e_k。评估拟合优度的基础是统计量的分布：

$$\sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i} \tag{3.23}$$

当 $N\to\infty$ 时，此统计量的分布接近自由度为（$f=k-1$）的卡方（χ^2）分布，χ^2 分布有以下的概率密度函数：

$$f(x) = \frac{1}{2^{\frac{f}{2}}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} d^{\frac{x}{2}-1} \quad (x>0) \tag{3.24}$$

式中，f 是自由度，$\Gamma\left(\frac{f}{2}\right)$是伽马函数。$\chi^2$ 分布的累积概率函数在任何一本统计书中都可以找到。

如果理论分布的参数未知，且是从数据中估计得来，如果每个需要被估计的未知参数的自由度都减 1，先前的分布仍然有效。

在此基础上，如果假设分布得到以下结果：

$$\sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i} < C_{1-\alpha,f} \tag{3.25}$$

那么，对于显著水平 α，假设理论分布是可接受的 [也就是所谓的虚假设（Null Hypothesis）H_0：$F(x)=F_0(x)$是可接受的]。如果不等式（3.25）不成立，则接受假设 H_0:$F(x)\neq F_0(x)$。在不等式（3.25）中，$C_{1-\alpha,f}$是与累积概率（$1-\alpha$）对应的χ^2 的值。

在使用卡方拟合优度检验时，我们推荐至少使用 5 个区间（$k>5$），每个区间内至少有 5 组观测数据（$e_i>5$），以获得满意结果。执行卡方检验的步骤如下：

① 把数据范围划分成包括第一个和最后一个无限区间的区间（区间的数量 >5），计数 n_i = 每个区间内的观测次数。

② 为假设理论分布 $F_0(x)$ 估计参数，并计算每个区间内数据的理论量 e_i 如下：

$$e_i = [F_0(x+\Delta x) - F_0(x)][\text{样本量}] \tag{3.26}$$

③ 计算公式（3.23）。

④ 选择一个特定的显著性水平 α（通常，$1-\alpha=90\%$ 或者 95%），确定χ^2 分布的自由度：

$$f=k-1-[F_0(x)\text{的参数量}] \tag{3.27}$$

从表中查出 $C_{1-\alpha,f}$，并把它和公式（3.23）的计算结果作比较。如果不等式（3.25）成立，那么假设理论分布函数 $F_0(x)$ 是可接受的。

案例 3.3

在某微电子装置的寿命试验中，138个失效时间布在7个时间区间内（$k=7$）。每个区间内的失效次数在表3.1中列出。用显著性水平为 $\alpha=5\%$ 的 χ^2 检验确定正态分布和指数分布的相对拟合优度。

表3.1 寿命分布的卡方检验（案例3.3）

区间序号	区间范围 $t_{i-1}-t_i$/h	观测到频率	理论频率 e_i		$\sum_i \frac{(n_i-e_i)^2}{e_i^2}$	
			正态分布	指数分布	正态分布	指数分布
1	2000 ~ 3000	1	0.184	54.298	0.042	52.317
2	3000 ~ 3500	11	7.314	11.659	1.900	52.354
3	3500 ~ 4000	24	15.180	6.856	7.027	95.227
4	4000 ~ 4500	33	26.496	6.203	8.621	210.980
5	4500 ~ 5000	31	32.430	5.613	8.684	325.800
6	5000 ~ 5500	22	28.014	5.079	9.090	382.180
7	5500 ~ 6500	16	22.080	6.727	9.341	394.965

由数据可得出均值和方差的估计量 $\overline{T}=4545.3\text{h}$、$S^2=829.2\text{h}^2$，计算公式如下：

$$\overline{T}=\frac{\sum_{i=1}^{k}\frac{1}{2}(t_i+t_{i-1})n_i}{\sum n_i} \tag{3.28}$$

$$S=\sqrt{\frac{\sum_{i=1}^{k}\left(\frac{1}{2}(t_i+t_{i-1})-\overline{T}\right)^2 n_i}{\sum_{i=1}^{k}n_i-1}} \tag{3.29}$$

那么，正态分布的累积分布函数为

$$F_{T,N}(t)=\Phi\left(\frac{t-\overline{T}}{S}\right)=\Phi\left(\frac{t-4545.3}{29.2}\right) \tag{3.30}$$

指数分布的参数估计为 $\lambda=1/\overline{T}=0.00022$，那么，指数分布的累积分布函数如下：

$$F_{T,E}(t)=1-\mathrm{e}^{\lambda t}=1-\mathrm{e}^{-0.00022t} \tag{3.31}$$

每个区间内数据的理论量 e_i 可以通过下面公式的计算得到：

$$e_{i,N}=138[F_{T,N}(t_i)-F_{T,N}(t_{i-1})] \tag{3.32}$$

$$e_{i,E}=138[F_{T,E}(t_i)-F_{T,E}(t_{i-1})] \tag{3.33}$$

正态分布的自由度 $f=7-1-2=4$，指数分布的自由度 $f=7-1-1=5$。在5%显著性水平上，正态分布情况下，$C_{95\%,1}=9.49$；指数分布情况下，$C_{95\%,5}=11.1$。通过对比这些值和表中第6列和第7列中 $\sum(n_i-e_i)^2/e_i^2$ 的值可以看出：在显著性水平为

5% 的拟合优度检验中，正态分布是可接受的，指数分布是不可接受的。指数分布并不适合于这些数据，因为其均值等于标准差。在本例中，样本均值为 150，大于样本标准差。

2. Kolmogorov-Smirnov（K-S）检验

另外一种广泛应用的拟合优度检验是 K-S 检验。其基本程序是把经验（或抽样）分布函数（Empirical Distribution Function，E. D. F.）和假定理论分布函数相比较，如果最大的偏差比从已知样本大小得到的期望值大，那么这个假设理论模型就是不可接受的。

设由随机变量的 n 次观测值得到的一个未检验样本（Uncensored Sample），对数据进行重新排列，使之单调递增：$X_1 < X_2 < \cdots < X_n$。使用排序后的样本数据，经验分布函数 $S_n(x)$ 的定义如下：

$$S_n(X) = \begin{cases} 0 & -\infty < X < X_1 \\ \dfrac{i}{n} & X_i \leqslant X < X_{i+1} \qquad i = 1, \cdots, n-1 \\ 1 & X_n \leqslant X < \infty \end{cases} \tag{3.34}$$

式中，X_1，X_2，…，X_n 是有序样本数据（有序统计量）的值。图 3.2 是 $S_n(x)$ 的曲线和理论分布函数 $F_0(x)$ 的曲线。

大数定律说明，经验分布函数是与理论分布函数一致的估计量。

在 K-S 检验中，整个 x 范围内 $S_n(x)$ 和 $F_0(x)$ 之间的最大偏差（检验统计量）用来衡量理论模型和经验分布函数之间的差异。$S_n(x)$ 和 $F_0(x)$ 之间的最大偏差记作

$$D_n = \max_x |F_0(x) - S_n(x)| \tag{3.35}$$

如果虚假设为真，那么对于任一可能连续的 $F_0(x)$，D_n 的概率分布都是一样的。因此，D_n 是一个其分布仅与样本量 n 相关的随机变量。

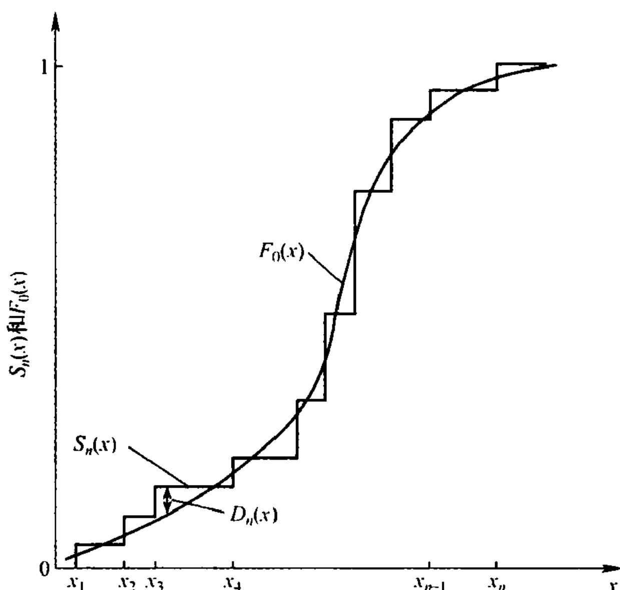


图 3.2　经验分布函数和理论分布函数

对于特定的显著性水平 α，K－S 检验对比观测到的最大偏差和临界值 D_n^α，即

$$P(D_n \leqslant D_n^\alpha) = 1 - \alpha \tag{3.36}$$

在各类统计书籍中，不同显著性水平 α 上的临界值 D_n^α 都是以列表的形式给出的（参见参考文献）。如果得到的 D_n 小于临界值 D_n^α，那么假设分布是可接受的。K-S 检验的实施步骤如下：

① 对每个样本数据项，用公式 3.34 计算 $S_n(x_i)$（$i = 1$，…，n）。

② 用另一样本估计假设理论分布 $F_0(x)$ 的参数，并从假设理论分布函数计算$F_0(x_i)$。

③ 为每个样本数据计算 $S_n(x_i)$ 和 $F_0(x_i)$ 的偏差，根据公式（3.35）确定偏差的最大值。

④ 选择一个特定的显著性水平 α（通常，所有检测都选择 $1-\alpha=90\%\sim95\%$），并从相应统计表中查出临界值 D_n^α [Beyer，1968]。

⑤ 比较 D_n 和 D_n^α，如果 D_n 小于 D_n^α，则假设理论分布 $F_0(x)$ 是可接受的。

案例 3.4

对砷化镓晶圆的断裂模量进行试验，表3.2中列出了11个晶圆的试验结果，用K-S检验判断这些数据在显著性水平 $\alpha=5\%$ 时是否为正态分布。

表3.2 断裂模量的K-S检验（案例3.4）

序 号	断裂模量 X_i/MPa	$S_k(x_i)$	$F_n(x_i)$	$D_n=\lvert F_n(x_i)-S_n(x_i)\rvert$
1	67.38	0.091	0.200	0.109
2	69.96	0.182	0.249	0.067
3	71.00	0.273	0.269	0.004
4	73.22	0.364	0.318	0.046
5	74.75	0.455	0.352	0.103
6	75.67	0.545	0.374	0.171
7	80.37	0.636	0.488	0.148
8	81.64	0.727	0.518	0.209
9	84.23	0.818	0.583	0.235
10	85.50	0.909	0.614	0.295
11	125.71	1.000	0.997	0.003

表3.2中的数据显示，样本的均值为 $\overline{X}=80.86$MPa，样本的标准差为 $\sigma_x=16.02$MPa。计算表3.2中第三栏的 $S_n(x)=i/n$，另外，对每个样本元素计算

$$F_0(X_i)=\Phi\left(\frac{X_i-80.86}{16.02}\right) \tag{3.37}$$

和$\lvert F_0(X_i)-S_n(X_i)\rvert$，结果在表3.2中列出。从这些结果可以看出，$x=85.5$MPa时，最大的绝对偏差 $D_n=0.295$。此案例有11个试验数据点，因此，显著性水平为5%时，D_n^α 的临界值为 $D_n^{0.05}=0.40$。因为最大偏差0.295比 $D_n^{0.05}$ 小，所以在显著性水平为5%时，把砷化镓晶圆的断裂模量试验的结果假设为正态分布是可接受的。

3. 样本比较

在一些试验条件下，我们必须对两个或者更多样本做出比较。例如在两种复杂应力条件下比较同一装置的两个样本失效次数，试验的目的是判断装置在其中一种应力条件下的可靠度是否与另一应力条件下的可靠度相同，或许还要计算失效时间（Time To

Failure, TTF）分布和可靠度的具体值。这类问题和非参数统计检验（Nonparametric Statistical Test, Distribution Free Statistical Test）相关。对两个样本进行的 Mann-Whitney 检验（Wilcoxon）就是一种非参数检验。

假设观测量（X_i, …, X_m）和（Y_i, …, Y_n）的两个独立样本分别来自分布函数为 $F_X(x)$ 和 $F_Y(y)$ 的连续分布，则：

① 第一个样本的每个测量值和第二个样本的每个测量值有相同分布，即$F_X(x) = F_Y(y)$（虚假设，H_0）。

② 存在一个常量 Θ，使得每个随机变量（$Y_i - \Theta$）和 X_t 有相同的分布，即$F_X(x) = F_Y(x-\Theta)$（备选假设，H_1）。

把这些假设记作：

$$H_0: F_X(x) = F_Y(x)(-\infty < x < +\infty),$$
$$H_1: F_X(x) = F_Y(x-\Theta), \Theta \neq 0(-\infty < x < +\infty)。$$

备选假设也可以记作 $\Theta > 0$ 或者 $\Theta < 0$。

如果把这两个样本合并为一个样本，则合并后样本的有序统计量为 Z_1, Z_2, …, Z_{m+n}，且 $Z_1 < Z_2 < \cdots < Z_{m+n}$。样本元素的等级和它在当前序列中所处的位置 Z_i一致，因此最小样本元素的等级是1，第二小样本元素的等级是2，以此类推。

把合并样本 Z 中第一个样本 X_t的所有元素的等级记为 $Z_i(i=1, 2, \cdots, m+n)$，将这些等级的和记为 S。因为在合并样本中，所有等级的平均数是（$1+m+n)/2$，很明显，如果 H_0 为真，那么

$$E(S) = \frac{m(m+n+1)}{2} \tag{3.38}$$

可以得到

$$\mathrm{Var}(S) = \frac{mn(m+n+1)}{12} \tag{3.39}$$

此外，这个例子中的 m 和 n 都大于8，S 是近似正态分布，其均值和方差与合并前相同。当 m 和 n 都小于8时，就要用查特定的数据表 [Dixon and Massey, 1969]。

对于可靠性试验，我们最关心的是验证假设 $H_0: F_X(x) = F_Y(y), -\infty < x < +\infty$相对于备选假设 $H_1: F_X(x) = F_Y(x-\Theta), \Theta > 0$ 的有效性。如果 x 和 y 都是失效时间，那么此假设等同于假设（可靠度函数表示）$R_X(x) > R_Y(y)$，也就是等同于假设：来自 $F_X(x)$的第一个样本的项比来自 $F_Y(y)$ 的第二个样本的项更可靠。

如果 $S > C(m, n, \alpha)$，否决假设 H_0，其中，α 是当 $S > C(m, n, \alpha)$ 成立时否决假设 H_0 的概率（显著性水平）。对于小样本（m 和 n 都小于8）$C(m, n, \alpha)$ 的值，以表格的形式列出 [Dixon and Massey, 1969]；对于较大样本，如果 $C > z_\alpha$，拒绝假设，其中：

$$C = \frac{S - E(S)}{[\mathrm{Var}(S)]^{1/2}} \tag{3.40}$$

z_α 是标准正态分布 N（0, 1）的第 100α 个百分位数。

案例 3.5

假定有不同应力条件下某装置失效时间的两个样本：

样本1（h）：90，367，470，572，1307，1345，1392，1603，2152，2858；$m=10$。

样本2（h）：37，150，154，319，373，433，538，571，751，1180；$n=10$。

虚假设：这个装置在两种应力条件下具有相同的可靠度；备选假设：此装置在第二种应力条件下的可靠度较差。假设检验的显著性水平为5%（即 $\alpha=0.05$），合并样本的有序排列为 37（1），90（2*），150（3），154（4），319（5），367（6*），373（7），433（8），470（9*），538（10），577（11），572（12*），751（13），1180（14），1307（15*），1345（16*），1392（17*），1603（18*），2152（19*），2858（20*）。

其中，带有星标的等级表示合并样本 Z 中第一个样本 X_1 的元素的等级。这些等级的和为

$$S=134, E(S)=105, \mathrm{Var}S=2100/12=175, (\mathrm{Var}S)^{1/2}=13.23$$

因此，$C=(134-105)/13.23=2.19$。

从正态分布表［Beyer，1968］可查出 $C_{0.05}=1.64$。因此，拒绝假设 H_0；第二种应力条件下设备的可靠度较差。

3.4　可靠性回归模型的拟合

上述大部分内容讨论的是单一随机变量的情况。但是可靠性问题往往需要了解几个随机变量之间的概率关系。例如，装置的失效时间可能取决于外施电压、环境温度、湿度等。失效时间可以看做是一个随机变量 Y，它是变量 x_1（电压）、x_2（温度）和 x_3（湿度）的函数。这种函数必然涵盖多种不确定性，因此这类模型被广泛采用。

3.4.1　Gauss-Markov 理论和线性回归

1. 回归分析

在回归分析（Regression Analysis）中，Y 被看作是因变量（Dependent Variable）或者响应值（Response），x_1、x_2 和 x_3 被看作是自变量（Independent Variable）或者因数（Factor）。通常情况下，自变量 x_1，…，x_k 可能是随机变量或者是非随机变量，其值是已知的或由试验者自己选择。任何已知 x_1，…，x_k 值的 Y 条件期望［$E(Y|(x_1, \cdots, x_k))$］称为 Y 在 x_1，…，x_k 的回归（Regression）。

如果 Y 回归是自变量 x_1，…，x_k 的一个线性函数，那么

$$E(Y|x_1,\cdots,x_k)=\beta_0+\beta_1x_1+\cdots+\beta_kx_k \tag{3.41}$$

系数 β_0，β_1，…，β_k 称为参数的回归系数。在 Y 期望是一个非随机变量的情况下，关系式（3.41）是确定的。与随机变量 Y 对应的回归模型记作：

$$Y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+\varepsilon \tag{3.42}$$

式中，ε 称为随机误差（Random Error），它服从一个均值为 $E(\varepsilon)=0$，有限方差为 σ^2 的分布。如果 ε 为正态分布，就可以用它来处理正态回归（Normal Regression）。

简单线性回归（Simple Linear Regression），有简单确定性关系的回归模型称为简单

线性回归，为

$$Y = \beta_0 + \beta_1 x \tag{3.43}$$

假设有 n 对测量值 (x_1, y_1)，…，(x_n, y_n)，如图 3.3 所示，随着 x 值的增长，Y 的总趋势是增长的。另外，假设对于任何已知变量 x，因变量 Y 和自变量 x 有以下关系：

$$Y = \beta_0 + \beta_1 x + \varepsilon \tag{3.44}$$

式中，ε 服从均值为 0，方差为 σ^2 的正态分布。对于已知的 x，随机变量 Y 服从均值为 $\beta_0 + \beta_1 x$，方差为 σ^2 的正态分布。因此，回归模型是随机变量 Y 的位置变形。也就是说，随机变量 Y 由非随机变量 $\beta_0 + \beta_1 x$ 加上随机变量 ε 形成，还可以推出，对于任意已知 $(x_1, \cdots, x_n)$ 值，随机变量 $(Y_1, \cdots, Y_n)$ 都是独立的。对于先前的 n 对测量值，$(y_1, \cdots, y_n)$ 的联合概率密度函数计算为

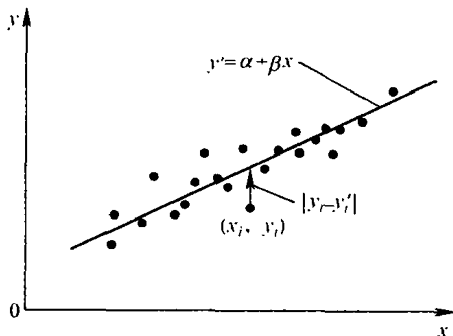


图 3.3　线性回归的拟合

$$f_n(y \mid x, \beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right] \tag{3.45}$$

函数（3.45）是参数 β_0 和 β_1 的似然函数。β_0 和 β_1 函数的最大化问题可简化为求 β_0 与 β_1 平方和的最小值：

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \tag{3.46}$$

因此，参数 β_0 和 β_1 的极大似然估计，采用的是最小二乘估计法。Gauss-Markov 理论定义了最小二乘估计的性质，后面将对此进行讨论。

β_0 和 β_1 的值以及 $S(\beta_0, \beta_1)$ 的最小值由式（3.47）确定：

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0 \quad \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0 \tag{3.47}$$

以上方程的解就是参数 β_0 和 β_1 的最小二乘估计（记作 $\hat{\beta}_0$ 和 $\hat{\beta}_1$），即

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) y_i}{\sum_{i=n}^{n}(x_i - \bar{x})^2} \tag{3.48}$$

式中，

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.49}$$

注意：以上估计值是测量值 y_i 的线性函数。

因变量方差 σ^2 的估计为

$$S^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{(n-2)} \tag{3.50}$$

式中，

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{3.51}$$

它是通过因变量的回归模型预测而来的，$(n-2)$ 是自由度，2 是模型中被估参数

的数量。

我们可以证明估计量$\hat{\beta}_0$和$\hat{\beta}_1$是正态分布随机变量，其均值分别为β_0和β_1。$\hat{\beta}_0$和$\hat{\beta}_1$的联合分布是协方差（Covariance）为Cov（$\hat{\beta}_0$，$\hat{\beta}_1$）的二元正态分布（Bivariate Normal Distribution），协方差为

$$\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)=\frac{\bar{x}s^2}{\sum_{i=1}^{n}(x_i-\bar{x})} \tag{3.52}$$

遗憾的是，通常情况下的估计量$\hat{\beta}_0$和$\hat{\beta}_1$是相互联系的。为了避免这种情况，必须选择使得公式（3.52）中样本的均值x等于0的x_i值。此选择可作为一个简单的试验设计案例。

已经获得的估计量$\hat{\beta}_0$、$\hat{\beta}_1$和s^2，可用来构建以下置信区间：

β_0的（$1-\alpha$）双边置信区间为

$$\hat{\beta}_0 \pm t_{n-2;\alpha/2}(s^2)^{1/2}\left[\frac{1}{n}+\frac{(\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{1/2} \tag{3.53}$$

β_1的（$1-\alpha$）双边置信区间为

$$\hat{\beta}_1 \pm t_{n-2;\alpha/2}(s^2)^{1/2}\left[\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{1/2} \tag{3.54}$$

对于任何已知点x_0，Y的均值的（$1-\alpha$）双边置信区间为

$$\hat{y}(x_0) \pm t_{n-2;\alpha/2}(s^2)^{1/2}\left[\frac{1}{n}\ \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{1/2} \tag{3.55}$$

基于以上参数估计量的分布，我们可以对下面的假设进行检验：

① 令β_0^*为一个已知数。检验回归参数β_0等于β_0^*的假设（虚假设）相对于备选假设β_0不等于β_0^*的有效性，即

$$H_0:\beta_0=\beta_0^*$$
$$H_0:\beta_0\neq\beta_0^*$$

② 其他条件同上，检验以下假设：

$$H_0:\beta_1=\beta_1^*$$
$$H_0:\beta_1\neq\beta_1^*$$

③ 检验关于β_0和β_1的假设：

H_0：$\beta_0=\beta_0^*$ 且$\beta_1=\beta_1^*$

H_1：假设H_0不成立。

相关系数（Correlation Coefficient）ρ在回归分析中有额外的含义。假设随机变量Y和x服从二元正态分布，此时，对于已知x，Y的条件分布是一元正态分布（Univariate Normal Distribution），其方差由下式计算：

$$\sigma^2 = \sigma_Y^2(1 - \rho^2) \tag{3.56}$$

式中，σ_Y^2 是当 x 未知时 Y 的无条件方差（Unconditional Variance）（也是 Y 的方差）。从公式（3.56）可得

$$\rho^2 = \frac{\sigma_Y^2 - \sigma^2}{\sigma_Y^2} \tag{3.57}$$

此关系式有一个重要的含义。它意味着相关系数的平方值等于从已知 X 获得的 Y 的部分方差值。

2. Gauss-Markov 理论

设因变量的 n 个观测值为 Y_1，…，Y_n，假设期望 $E(Y_i)$ 由公式（3.41）计算为

$$\begin{aligned} &E(Y_i) = \beta_0 x_{0i} + \cdots + \beta_k x_{ki} \\ &i = 1, \cdots, n, n > k + 1 \end{aligned} \tag{3.58}$$

式中，x_{0i}，…，x_{ki}为自变量的已知值，从 Y_i 值的抽样而来（抽样规则为 $x_{0i} \equiv 1$）。因此每个测量值 Y_i 可记作：

$$\begin{aligned} &Y_i = \beta_0 x_{0i} + \cdots + \beta_k x_{ki} + \varepsilon_i \\ &i = 1, \cdots, n, n > k + 1 \end{aligned} \tag{3.59}$$

式中，ε_i 是随机无关偏差（Random Uncorrelated Error）$[\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0], E(\varepsilon_i) = 0, \mathrm{Var}(\varepsilon_i) = \sigma^2 (i, j = 1, \cdots, n)$。

这些假设构成了一般线性模型（General Linear Model）。需要注意的是：此处没有关于随机偏差的正态分布的假设。对于简单线性回归，β_0，…，β_k 的最小二乘估计值为 $\hat{\beta}_0$，…，$\hat{\beta}_k$，其平方和的最小值为

$$SS(\beta_0, \cdots, \beta_k) = \sum_{i=1}^{n} (Y_i - \beta_0 \beta_{0_i} - \cdots - \beta_k \beta_{k_i})^2 \tag{3.60}$$

在一般线性模型下，最小二乘估计是无偏估计，且所有的无偏估计都有最小的方差，无偏估计在因变量测量值中是线性估计。以上表述称作 Gauss-Markov 理论。

3. 多重线性回归

一般线性模型可以记作简单的矩阵形式，令 $Y = (Y_i, \cdots, Y_n)'$，$\beta = (\beta_0, \cdots, \beta_k)'$，$\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$，且

$$X = \begin{pmatrix} x_{01} & \cdots & x_{k1} \\ x_{02} & \cdots & x_{k2} \\ \vdots & \cdots & \vdots \\ x_{0n} & \cdots & x_{kn} \end{pmatrix} \tag{3.61}$$

把 A'记作所有矢量或矩阵 A 的变换。那么公式（3.59）变形为

$$Y = X\beta + \varepsilon \tag{3.62}$$

可以证明估计量 $\hat{\beta} =（\hat{\beta}_0, \cdots, \hat{\beta}_k）$ 的矢量计算为

$$\hat{\beta} = (X'X)^{-1} X'Y \tag{3.63}$$

通常情况下，系数 β 的估计量 $\hat{\beta}$ 是相互关联的。$\hat{\beta}$ 的协方差矩阵（Covariance Ma-

trix）为

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \tag{3.64}$$

有时，称矩阵 X 为试验的设计矩阵。此公式是最优试验设计的基本原则，因为几乎所有的最优试验设计都是用协方差矩阵表达的。例如设计矩阵是正交的，$X'X$ 是单位矩阵，那么所有的估计量 $\hat{\beta}$ 是独立的（无关联的）随机变量，且拥有相等的方差 σ^2。注意：任何一个试验的优化设计都建立在模型（3.58）的先验形式已知的基础之上。

所有这些表述都在一般线性模型范围内有效——即没有关于偏差服从正态分布的假设。另外，如果随机偏差服从正态分布，最小二乘估计是所有无偏估计中（包括那些 Y_t 的非线性函数）拥有最小方差的估计，这就是多重线性回归。此时，可以证明估计量 $\hat{\beta}$ 服从正态分布，所以我们可以构造不同的置信界限并检验一些假设。所要检验的大部分假设和简单线性回归使用的假设非常相似，但是也有一些假设拥有多元特征。例如试验者可以检验此模型的自变量 x_i 中的一个变量对因变量 Y 没有影响的假设，也就是检验假设：

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

在许多实际情况下，试验者也可能关心自变量在预测因变量时的重要性排序。例如对于一个给定的设备，试验者可能需要对影响可靠度（自变量）的因素（载荷、温度、湿度等）进行排序。对于任意自变量 $X_i (i=1, \cdots, k)$，检验前面提到的假设（$\beta_i=0$），这样并不能给出所关注因素的排序。在这种情况下，运用逐步回归法（Stepwise Regression Method）将很有效［Draper and Smith，1981］。

3.4.2　比例风险（PH）模型和加速寿命（AL）模型

可靠性模型通常被定义为装置的失效时间分布和应力（如载荷、循环率、温度、湿度和电压等）之间的关系，我们也可以认为它是失效时间的随机变量的一种确定性变形。寿命数据分析中有两种主要的时间转换：加速寿命（Accelerated Life，AL）模型和比例风险（Proportional Hazard，PH）模型。

1. 加速寿命（AL）模型

令 $F_1(t; z_1)$ 和 $F_2(t; z_2)$ 分别为某装置在恒定应力条件 z_1 和 z_2 下的失效时间累积分布函数，如果应力条件 z_2 比 z_1 更加苛刻，对于所有值为正的 t，有

$$F_2(t;z_2) > F_1(t;z_1) \tag{3.65}$$

此不等式意味着更苛刻的应力条件会加速失效。在不失一般性的情况下，我们可以为正态（运行）应力条件假设 $z=0$。如果把正态应力条件下的失效时间累积分布函数记作 $F_0(\times)$，那么，AL 时间变形用已知函数 $F(t; z)$ 和 $F_0(\times)$ 表示为［Cox and Oaks，1984］

$$F(t;z) = F_0[t\psi(z,\boldsymbol{A})] \tag{3.66}$$

式中，$\psi(z, \boldsymbol{A})$ 是把失效时间和应力因素 z 的一个矢量相关联的函数，$\boldsymbol{A}$ 是未知参数的一个矢量。通常，$\psi(z, \boldsymbol{A})$ 与递减的失效时间相对应。对于 $z=0$，假设 $\psi(z, \boldsymbol{A})$ 等

于 1。

公式（3.66）是一个比例变换。它表示应力的改变不会引起分布函数形状的改变，而仅仅改变它的比例。

对于分布函数 $F_1(t;\ z_1)$ 和 $F_2(t;\ z_2)$，如果 z_1 比 z_2 的严重度稍差，t_1 和 t_2 是 $F_1(t_1;\ z_1)=F_2(t_2;\ z_2)$ 时的失效时间，那么，存在一个函数 g（对于所有为正值的 t_1 和 t_2），使得 $t_1=g(t_2)$，因此

$$F_2(t_2;z_2)=F_1(g(t_2);z_1) \tag{3.67}$$

由于 $F_1(t;\ z_1)<F_2(t;\ z_2)$，$g(t)$ 是一个递增函数，$g(0)=0$ 并且

$$\lim_{t\to\infty} g(t)=\infty \tag{3.68}$$

函数 $g(t)$ 称为加速变换函数（Acceleration Transformation Function）或时间变换函数（Time Transformation Function）。公式（3.66）中的假设为：应力条件的改变不会引起分布函数形状的改变，而仅改变它的比例。可以用加速函数将 $g(t)$ 记为

$$g(t)=\psi(z,\boldsymbol{A})t \tag{3.69}$$

换种说法就是：公式（3.66）等价于线性时间加速函数（Time Acceleration Function）。通过公式（3.66）获得的失效时间的第 100 个百分位数 $t_p(z)$ 和风险率 $\lambda(z)$ 之间的关系为

$$t_p(z)=t_p^0/\psi(z,\boldsymbol{A}) \tag{3.70}$$

$$\lambda(t;z)=\psi(z,\boldsymbol{A})\lambda^0[t\psi(z,\boldsymbol{A})] \tag{3.71}$$

式中，t_p^0 和 λ^0 为正态应力条件 $z=0$ 下的第 100 个百分位数和故障率。

2. 比例风险（PH）模型

PH 模型的基本关系式类似于公式（3.66），即

$$F(t;z)=1-[1-F_0(t)]^{\psi(z,\boldsymbol{A})} \tag{3.72}$$

我们认为比例风险（Cox）模型是风险与故障率（Hazard Rate）的关系，可以从公式（3.72）中得到

$$\lambda(t;z)=\psi(z,\boldsymbol{A})\lambda^0(t) \tag{3.73}$$

通常，式中的 $\psi(z,\ \boldsymbol{A})$ 是一个对数线性函数。

PH 模型时间变换通常不再拥有累积函数分布的形状，函数 $\psi(z)$ 与加速函数的关系也不再简单。因此，PH 模型在可靠性应用中不像 AL 模型那样常见。

然而，我们可以证明［Cox and Oaks，1984］：PH 模型和 AL 模型仅对于 Weibull 分布是一致的。在可靠性应用中，AL 模型时间变换更加常用，而 PH 模型则在生物医学的寿命数据分析中有着广泛的应用。

3.4.3 恒定应力下的加速寿命回归

假设要根据恒定应力条件下的加速寿命试验预测可靠度，对于第 100 个百分位数 t_p，假设可靠度模型 $\eta(z,\ \boldsymbol{B})$ 是关于应力因素 z 和参数的未知矢量 $\boldsymbol{B}$ 的已知函数

$$t_p(z,\boldsymbol{B})=\eta(z,\boldsymbol{B}) \tag{3.74}$$

与公式（3.70）相关的可靠度模型为

$$\eta(z,\boldsymbol{B})=t_p^0/\psi(z,\boldsymbol{A}) \tag{3.75}$$

对于百分位数（包含中值），最常用的模型是对数线性模型（Log-Linear Model）。幂法则模型（Power Rule Model）和阿伦尼斯反应速率模型（Arrhenius Reaction Rate Model）都是这类模型。对于幂法则模型

$$t_{\mathrm{p}}(x) = a/x^{c}, c>0, x>0 \tag{3.76}$$

式中，x 是机械应力或电气应力。

对于阿伦尼斯反应速率模型

$$t_{\mathrm{p}}(T) = a\ \exp(E_{\mathrm{a}}/T) \tag{3.77}$$

式中，T 是热力学温度，E_{a} 是激活能（Activation Energy）。

将这两种模型组合，得到的模型为

$$t_{\mathrm{p}}(x,T) = a\ x^{-c}\exp(E_{\mathrm{a}}/T) \tag{3.78}$$

式中，a、E_{a} 和 c 都是需要估计的参数。

我们的目的是获得模型（3.74）中矢量 $\boldsymbol{B}$ 的一个估计量，根据不同应力条件下 z_1，…，z_k 的 AL 试验预测正态（已知）应力条件下可靠度的百分位数。其中，k 大于矢量 $\boldsymbol{B}$ 的维数，即 $k>\dim \boldsymbol{B}$。

另外，假设：

① 在所有应力条件下的 TTF（失效时间）分布是关于连续密度函数 $f(t, z)$ 的递增平均失效率（Increasing Failure Rate Average，IFRA）分布。

② 试验结果是第Ⅱ类检验合格样本，未检验失效时间的数量 $r_i(i=1, \cdots, k)$ 和样本大小 n_i 足够大，以至于可以把 t_{p} 估计为样本的百分位数 $\hat{t}_{\mathrm{p}}$：

$$\hat{t}_{\mathrm{p}} = \begin{cases} t_{(\lfloor n,p \rfloor)} & (n,p\ \text{不为整数}) \\ \text{区间}[t_{(n,p)}, t_{(n,p+1)}]\text{内的任意值} & (n,p\ \text{为整数}) \end{cases} \tag{3.79}$$

式中，$t_{(x)}$ 是失效时间（顺序统计量）；样本的大小允许使用估计量的渐近正态分布。基于前面的假设，公式（3.74）的模型可以记作

$$\hat{t}_{\mathrm{p}}(z,\boldsymbol{B}) = \eta(z,\boldsymbol{B})\varepsilon \tag{3.80}$$

其中，

$$\varepsilon \sim N\left\{1, \frac{p(1-p)}{n_i\eta^2(z,\boldsymbol{B})f^2[\eta(z,\boldsymbol{B})]}\right\} \tag{3.81}$$

式中，$N(a, b)$ 是均值为 a，方差为 b 的正态分布。

乘法模型（3.80）可以通过有偏正态分布模型的对数变换变成标准正态回归：

$$\begin{aligned} \ln \hat{t}_{\mathrm{p}}(z,\boldsymbol{B}) &= \ln \eta(z,\boldsymbol{B}) + \varepsilon_1 \\ \varepsilon_1 &\sim N(0,\sigma^2) \end{aligned} \tag{3.82}$$

式中，σ^2 是一个未知常量。此变换的基础是递增平均故障率（Increasing Failure Rate Average，IFRA）分布的属性和一个概率变换：令 $x \sim N(0, \sigma^2)$，且 $\sigma \ll 1$，那么，随机变量 $y=\ln(1+x)$ 是一个近似 x 的分布［或 $y \sim N(0, \sigma^2)$］。这额外地限制了模型（3.80）变为公式（3.82）的可能性，也就是强加在样本量 n_i 上的限制条件，即

$$[p\ln_i(1-p)\ln^2(1-p)]^{1/2} \ll 1 (i=1,\cdots,k) \tag{3.83}$$

那么，前面的预测问题就简化为对正态回归参数的估计，后续要用到点预测和区间预测。标准的回归试验设计技术可以应用于使用公式（3.81）的 AL 试验。

3.4.4　时间相关应力下的加速寿命回归

时间相关应力下的加速寿命试验，如步进应力试验（Step-Stress Test）和斜坡试验（Ramp Test）都是非常重要的。例如最常见的测试金属氧化物半导体集成电路中的二氧化硅薄膜的可靠度就是斜坡电压试验（Ramp-Voltage Test）。在这个试验中，氧化薄膜会被随时间线性递增的电压击穿。

令 $z(t)$ 为时间相关的应力矢量，且 $z(t)$ 可积。此时，公式（3.66）可以记作 [Cox and Oaks，1984]：

$$F(t;z) = F_0[\psi(t)] \tag{3.84}$$

$$\psi(t^{(z)}) = \int_0^{t^{(z)}} \psi[z(s), \boldsymbol{A}]\mathrm{d}s \tag{3.85}$$

式中，$t^{(z)}$ 是装置在应力条件 $z(t)$ 下的失效时间。

失效时间的第 100 个百分位数 $t_p[z(t)]$ 对应的关系式，可以从公式（3.84）获得，即

$$t_p^0 = \int_0^{t_p z(t)} \psi[z(s), \boldsymbol{A}]\mathrm{d}s \tag{3.86}$$

利用公式（3.74）和（3.75），以上关系式可以重新记作

$$1 = \int_0^{t_p[z(t)]} \frac{1}{t_p^0\{\psi[z(s), \boldsymbol{A}]\}^{-1}}\mathrm{d}s = \int_0^{t_p[z(t)]} \frac{1}{\eta[z(s), \boldsymbol{B}]}\mathrm{d}s \tag{3.87}$$

公式（3.82）是 Miner 准则的精确概率形式 [Miner，1945]，它广泛应用于断裂力学，用来计算不同应力下的累积损伤。因此，时间相关应力下的使用寿命加速试验和使用 Miner 准则一样，都存在适用性的问题。此外，机械损伤累积（Mechanical Damage Accumulation）和电击穿（Electrical Breakdown）之间还应该进行一次有效的模拟。

模型（3.80）的时间相关模拟（Time-Dependent Analog）为

$$t_p^0 = \int_0^{\hat{t}_p[z(t)]} \psi[z(s), \boldsymbol{A}]\mathrm{d}s \tag{3.88}$$

式中，$\hat{t}_p[z(t)]$是装置在应力条件 $z(t)$ 下的样本百分位数。

在这个例子中，对矢量 $\boldsymbol{A}$ 和 t_p^0 [或对可靠度模型（3.74）] 的估计，不能简化成和上述恒定应力中估计对数线性回归模型的参数问题。

设有 k 个不同的时间相关应力条件 $z_i(t)$，$i=1$，2，…，k [$k>$（$\dim\boldsymbol{A}$）+1]，试验结果（和前面的例子一样）是第Ⅱ类已检验样本（Censored Sample）并且未检验失效时间的数量和样本大小足够大，使得可以把 t_p 估计为样本的百分位数 t_p^0。在这种情况下，可靠性模型的参数估计（对于矢量 $\boldsymbol{A}$ 和 t_p^0）可以通过用最小二乘法求解以下积分方程得到：

$$t_p^0 = \int_{i=1,2,\cdots,k}^{\hat{t}_p[z_i(t)]} \psi[z_i(s), A]\mathrm{d}s \tag{3.89}$$

案例 3.6

假定用于一个陶瓷电容器的失效时间的第十个百分位数 $t_{0.1}$ 的模型（3.78）的形式为

$$t_{0.1}(U,T) = aU^{-c}\exp(E_a/T) \tag{3.90}$$

式中，U 是外施电压，T 是热力学温度。

假设使用一个时间步进应力 AL 试验：将步进应力和温度常量相结合作为加速应力因子。试验开始时，为试验样本施加一个特定低电压 U_0，试验时序时间为 Δt。电压的增量为 ΔU，经过 Δt 时间内，在 $U+\Delta U$ 下对样本进行试验：

$$U(t) = U_0 + \Delta U\, En(t/\Delta t) \tag{3.91}$$

式中，$En(x)$ 表示最接近但不大于 x 的整数。在 $p \geqslant 0.1$ 项失效时结束试验，那么，试验的结果就是在每一个电压—温度组合下的样本百分位数。表 3.3 给出了 $\Delta U = 10\text{V}$，$\Delta t = 24\text{h}$ 的试验方案和结果。

表 3.3 陶瓷电容器的试验结果

温度/K	电压 U_0/V	TTF 百分位数估计/h
398	100	347.9
358	150	1688.5
373	100	989.6
373	63	1078.6

对于此案例，积分方程（3.89）具有以下形式：

$$a = \int_0^{t_{0.1}} \exp(-E_a/T)[U(s_i)]^c\mathrm{d}s \quad (i = 1,2,3,4) \tag{3.92}$$

用前面的数据解这个方程，能够为公式（3.88）求得以下估计量：

$a = 2.227\times10^{-8}\text{h/V}^{1.885}$；$E_a = 1321\times10^4\text{K}$；$c = 1.885$。

3.5 结论

和上一章一样，本章介绍了读者需要了解的基本统计技术（点估计和区间估计、假设检验、基本回归等），同时还介绍了一些具体的可靠性试验技术（风险比例和加速寿命模型）。

参考文献

Beyer, W. 1968. Handbook of tables for probability and statistics. Boca Raton, FL: CRC

Press.

Cox, D. R., and D. Oaks. 1984. The analysis of survival data. London: Chapman & Hall.

Dixon, W. J., and F. J. Massey, Jr. 1969. Introduction to statistical analysis, 3rd ed. New York: McGraw-Hill.

Draper, N., and H. Smith. 1981. Applied regression analysis. New York: John Wiley & Sons.

Miner, M. A. 1945. Cumulative damage in fatigue. Journal of Applied Mechanics 12: A159-A164.

第4章　产品可靠性分析的实用概率分布

4.1　引言

用于可靠性工程中的数据通常从零件、材料、制造中和制造后的试验、现场试用产品、保修返回等方面收集而来。如果用概率分布对这些数据进行建模，那么所用概率分布的性质可用来为产品设计、制造以及可靠性估计等制定决策。

本章将首先介绍离散型（Discrete）概率分布和连续型（Continuous）概率分布以及它们的主要特点，然后介绍常用于可靠性建模及故障率评估的两种离散型分布（二项分布和Poisson分布）和四种连续型分布（Weibull分布、指数分布、正态分布和对数正态分布）。

4.2　离散型分布

如果离散随机变量x等于大量离散值（x_0，x_1，x_2，$\cdots$，x_n）中任何的一个，那么对于$x=x_i$，就存在概率$f(x_i)$

$$f(x_i)P\{x=x_i\} \tag{4.1}$$

公式（4.1）中，$f(x_i)$称为概率质量函数（Probability Mass Function，PMF）[⊖]，累积分布函数（Cumulative Distribution Function）记作：

$$F(x_i)=P\{x\leqslant x_i\} \tag{4.2}$$

离散随机变量的均值μ和方差σ^2可用概率质量函数表示为

$$\mu = \sum_i x_i f(x_i) \tag{4.3}$$

$$\sigma^2 = \sum_i (x_i-\mu)^2 f(x_i) = \sum_i x_i^2 f(x_i) - \mu^2 \tag{4.4}$$

二项分布和Poisson分布是可靠性工程师常用的两种离散分布，它们常用于开发产品的抽样和验收计划。在基于产品的零件（材料）可靠性来评估产品可靠性的过程中，它们也具有一定作用。

4.2.1　二项分布

二项分布是一种离散型概率分布，它适用于所有试验或测试只有两种互斥结果的情况。例如在掷骰子时，特定点数出现（成功）的概率是1/6，不出现（失败）的概率是5/6。这种只有“成功”和“失败”两种可能输出结果的随机试验称为伯努利试验

⊖　离散型概率函数被称为概率质量函数，而连续型概率函数被称为概率密度函数。对于通用术语中的概率函数而言，“概率密度函数”可同时用于表述离散型和连续型概率函数。

(Bernoulli Trial)。当然，“成功”和“失败”的定义是由试验本身来决定的。在一些试验中，成功是指结果而不是某个特定值的概率。

对于二项分布，在 m 次试验中成功 k 次的概率，其概率质量函数 $f(x)$ 为

$$f(k)=\binom{m}{k}p^{k}q^{m-k}(0\leqslant p\leqslant 1,q=1-p,k=0,1,2,\cdots,m) \tag{4.5}$$

式中　p——定义成功的概率；

q 或 $1-p$——失败的概率；

m——独立试验的次数；

k——m 次试验中成功的次数，其组合公式为

$$\binom{m}{k}\equiv C_k^m=\frac{m!}{k!(m-k)!} \tag{4.6}$$

式中，! 是阶乘符号，由于 $(p+q)$ 等于 1，两边同求 j 次幂得到

$$(p+q)^j=1 \tag{4.7}$$

公式 (4.7) 左侧的二项式展开是 j 次成功的概率。例如对于 3 次试验，每次试验有相同的成功概率 p 和失败概率 q，公式 (4.7) 变为

$$(p+q)^3=p^3+3p^2q+3pq^2+q^3=1 \tag{4.8}$$

它的基础是一般方程：

$$\sum_{k=0}^{m}f(k)=F(m)=p\{k\leqslant m\}=(p+q)^m \tag{4.9}$$

$(p+q)^3$ 展开的四个子项，分别是成功 3 次、成功 2 次、成功 1 次和 1 次都不成功的概率值。那么，对于 $m=3$，成功概率 $=p$，$f(3)=p^3, f(2)=3p^2q, f(1)=3pq^2, f(0)=q^3$。

当产品有不同的成功和失败概率的时候，二项式展开同样有效。此时，二项式展开公式为

$$\prod_{i=1}^{m}(p_i+q_i)=1 \tag{4.10}$$

式中，i 表示一个由 m 个部件构成的系统的第 i 个部件。对于一个由三个不同部件的系统，二项式展开的形式为

$$\begin{aligned}(p_1+q_1)(p_2+q_2)(p_3+q_3)&=p_1p_2p_3+(p_1p_2q_3+p_1q_2p_3+q_1p_2q_3)+\\&\quad(p_1q_2q_3+q_1p_2q_3+q_1q_2p_3)+q_1q_2q_3=1\end{aligned} \tag{4.11}$$

公式右边的第一项是三次试验都成功的概率；第二项是任意两次试验成功的概率；第三项（括号内）是任意一次试验成功的概率；最后一项是所有试验都失败的概率。

二项分布 $F(k)$ 的累积分布函数是在 m 次试验中成功 k 次或者更少次的概率。它用离散型 PMF 定义为

$$F(k)=\sum_{i=0}^{k}(i) \tag{4.12}$$

或者用二项分布型 PMF 定义为

$$F(k) = \sum_{i=0}^{k} \binom{m}{i} p^i q^{(m-i)} \tag{4.13}$$

对于二项分布，其均值μ为

$$\mu = mp \tag{4.14}$$

方差为

$$\sigma^2 = mp(1-p) \tag{4.15}$$

案例　4.1

工程师从一大批电容中选出4个电容，在这批电容里有10%的次品，所选择4个电容符合下列条件的概率是多少？

(a) 有0个次品；

(b) 只有1个次品；

(c) 有2个次品；

(d) 2个或更少的次品。

解：此处，我们把成功定义为“选择一定数量的好电容”。所以，$p=0.1$，$q=0.9$，$m=4$。使用公式（4.5）和（4.6），$f(4)$ 是4个都是好电容（也就是没有次品）的概率，即

$$f(4) = \binom{4}{4}(0.9)^4(0.1)^0 = 0.6561$$

另外，四次试验有相同的p和q。

另外一种方法是定义成功为“选择一定数量的次品电容”，那么$p=0.1$，$q=0.9$。在这个情况下，$f(0)$ 是选择4个电容而不出现次品的概率，为

(a) $f(0) = \binom{4}{0}(0.1)^0(0.9)^4 = 0.6561$

继续使用第二种方法，问题（b）、（c）和（d）的解为

(b) $f(1) = \binom{4}{1}(0.1)^1(0.9)^3 = 0.2916$

(c) $f(2) = \binom{4}{2}(0.1)^2(0.9)^2 = 0.0486$

(d) $F(2) = f(0) + f(1) + f(2) = 0.9963$

案例　4.2

假设某产品在一种给定试验中的失效概率为0.1，如果有10个被测产品，

(a) 试验中预期失效次数为多少？

(b) 失效次数的方差是多少？

(c) 产品不发生失效的概率是多少？

(d) 2个或更多产品失效的概率是多少？

解：此处，$m=10$，$p=0.1$，则

(a) 预期失效次数的均值$\mu = mp = (10 \times 0.1) = 1$。

(b) 方差$\sigma^2 = mp(1-p) = [10 \times 0.1 \times (1-0.1)] = 0.9$。

（c）不出现失效的概率即概率质量函数中 $k=0$ 的值，为

$$f(0)=\binom{10}{0}\times 0.1^{0}\times(1-0.1)^{10}=0.349$$

（d）有 2 个或更多产品失效的概率等同于 1 减去没有或只有 1 个产品失效的概率

Pr(2 个或更多产品失效) $=\{1-[f(0)+f(1)]\}$

$=[1-0.394-\{10\times 0.1\times(1-0.1)^{9}]=0.264$

案例　4.3

某电子自动控制模块包括 3 个相同的、并行的微处理器。这些微处理器相互独立，其失效也相互独立。模块要正常工作，至少需要两个微处理器正常运行。每个微处理器在保修期内不失效的概率为 0.95，计算控制模块在保修期内的失效概率。

解：当两个或更多微处理器失效时，模块失效。也就是说，当只有一个或没有微处理器工作时，模块失效。那么模块在保修期内的失效概率为

$$Pr(\text{模块在保修期内失效})=[f(0)+f(1)]$$

式中，$m=3$ 是参与测试元件的数量，$k=0$ 或 1 是工作元件的总数量，$p=0.95$，$q=0.05$。因此：

$$Pr(\text{模块在保修期内失效})=\{(0.05)^{3}+[3\times 0.95\times(0.05)^{2}]\}=0.00725$$

$f(1)$=1个元件工作，2个元件失效 ⟩ 失效
$f(0)$=0 个元件工作，3个元件失效

$f(1)$=2个元件工作，1个元件失效 ⟩ 正常工作
$f(0)$=3 个元件工作，0个元件失效

Excel 中的二项分布函数：

① BINOMDIST（Number_ s，Trials，Probability_ s，Cumulative）返回二项式概率分布 $=f(k)$ 或 $F(k)$，其中：

Number_ s——在 k 次试验中未失效的元件数量；

Trials——指独立试验的数量 m；

Probability_ s——每次试验中不出现失效的概率 p；

Cumulative——确定函数形式的一个逻辑值［PMF（TRUE）或 CDF（FALSE)］。

② CRITBINOM（Trials，Probability_ s，Alpha）——累计值大于或等于一个标准值的二项分布返回 k 的最小值，其中：

Trials——指伯努利试验的数量 m；

Probability_ s——每次试验中不出现失效的概率 p；

Alpha——标准值，工作人员根据要解决的问题选择此项。

4.2.2　Poisson 分布

在成功的概率 p 非常低，且样本试验的次数（或者伯努利试验的次数）非常大的情况下，估计二项式的系数非常麻烦。此时，Poisson 分布可起到较好的作用。

Poisson 分布的 PMF 独立于试验的次数，记作

$$f(k)=\frac{\mu^k}{k!}e^{-\mu}(k=0,1,2,\cdots,n) \tag{4.16}$$

式中，μ 是均值也是方差。

对于 m 次伯努利试验的 Poisson 分布，其每次试验成功的概率为 p，均值和方差分别为

$$\mu=mp \tag{4.17}$$

$$\sigma^2=mp \tag{4.18}$$

Poisson 分布在工业和质量工程方面有广泛的应用，也是绘制控制图的基础。它适用于各种应用情况，例如确定生产环境中的污染颗粒、电力中断的次数和聚合物的缺陷等。

案例 4.4

用 Poisson 分布近似求解案例 4.2。

解：预期失效次数等于均值 $\mu=(10)(0.1)=1$，方差也等于 1。

不出现失效的概率与 $k=0$ 时的 PMF 相同，为

$$f(0)=e^{-\mu}=e^{-1}=0.3678$$

2 个或者更多产品失效的概率等于 1 减去没有或只有 1 个产品失效的概率，即

$$\begin{aligned}Pr(2\text{ 个或更多产品失效})&=[1-\{f(0)+f(1)\}]\\&=[1-\{0.3678+e^{-1}\}]\\&=0.2642\end{aligned}$$

注意此结果与案例 4.2 的结果的不同之处。

Excel 中的 Poisson 分布函数：POISSON（x，Mean，Cumulative）

x——事件的数量。

Mean——期望数值。

Cumulative——确定概率分布返回形式的一个逻辑值[PMF(TRUE)或 CDF(FALSE)]。

4.2.3 其他离散分布

其他一些在可靠性分析中可以使用的离散分布包括：几何分布、负二项分布（Negative Binomial Distribution）和超几何分布（Hypergeometric Distribution）。这些分布常用于一些特殊情况或二项分布受限的情况。

在几何分布中，伯努利试验持续进行直到第一次成功出现为止。几何分布有“记忆障碍”的特性，这意味着我们可以从任何一次试验开始统计试验次数，并不影响任何潜在的分布。在这一点上，几何分布和我们稍后将提到的连续指数分布（Continuous Exponential Distribution）有些相似。

在负二项分布中（一个广义的几何分布），伯努利试验持续进行直到一定数量的成功出现为止。它与二项分布的不同在于：成功的次数是既定的，而试验的次数是随机的。

在超几何分布中，替代测试在包含有多个不可替换产品或缺陷的样本中进行。超几何分布与二项分布的不同在于：它的总体是有限的，且从总体得来的样本不可替换。

4.3 连续型分布

如果把随机变量 x 的范围扩展到实数区间（有限或者无限），那么 x 就是一个连续

随机变量。累积分布函数的表达式为

$$F(x_i)=P\{x\leqslant x_i\} \tag{4.19}$$

$f(x)$ 是概率密度函数（等同于离散分布的概率质量函数，PDF），其表达式为

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}F(x) \tag{4.20}$$

由此可以得到

$$F(x)=\int_{-\infty}^{x}f(x)\mathrm{d}x \tag{4.21}$$

连续随机变量的均值 μ 和方差 σ^2 可以依据概率密度函数，在区间（$-\infty$，$+\infty$）上定义为

$$\mu=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x \tag{4.22}$$

$$\sigma^2=\int_{-\infty}^{+\infty}(x-\mu)^2f(x)\mathrm{d}x=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x-\mu^2 \tag{4.23}$$

如果 $f(x)$ 是失效概率密度函数［参见公式（2.4）］，当随机变量 x 表示的时间 $t\geqslant 0$ 时，那么，我们可以认为 $F(x)$ 是不可靠度 $Q(x)$ 的函数。因此，公式（4.21）等同于公式（4.5），公式（4.22）等同于公式（4.17）。

案例　4.5

某装置的 PDF（概率密度函数）是失效时间的函数，表达式为 $f(t)=\frac{1}{16}te^{-t/4}$。式中，$t$ 的单位为年，且 $t>0$。

（a）装置在第一年内失效的概率是多少？

（b）最少 5 年后才失效的概率是多少？

（c）如果不超过 5% 的装置在质保期内需要维修，那么装置质保期最长为几个月？

解： 对于已知 PDF，CDF 为

$$F(t)=\frac{1}{16}\int_0^t te^{-t/4}\mathrm{d}t=1-\left(\frac{t}{4}+1\right)e^{-t/4}$$

（a）在第一年内失效的概率为 $F(1)=0.0265$。

（b）至少 5 年后才失效的概率为 $[1-F(5)]=[1-0.3554]=0.6446$。

（c）对于问题所要求情况，$F(t_0)$ 必须小于 0.05，t_0 为保质期。从前面的结论我们可以得出：质保时间必须大于 1 年，而且 $F(2)$ 等于 0.09，因此，保修期应为 1～2 年。经过试验与误差估计，我们发现不超过 5% 的质保服务，$t_0=1.42$ 年。因此，保质期应设定为不超过 17 个月。

4.3.1　Weibull 分布

Weibull 分布是 Walloddi Weibull 在 1939 年提出的一个连续型分布（他也发明了球轴承和电锤）。Weibull 分布在可靠性分析中有广泛的应用，因为它可以对故障率曲线的广泛多样性进行建模。在特殊或者有一定限制条件的情况下，这种分布近似于其他分

布。Weibull分布也可应用于许多工程产品的寿命分布、材料强度和质保分析。

三参数Weibull概率分布函数的概率密度函数为

$$f(t)=\beta\eta^{-\beta}(t-\gamma)^{\beta-1}e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{4.24}$$

式中，$\beta>0$是形状参数，$\eta>0$是比例参数，γ是位置或时间延迟参数。可靠度函数的表达式为

$$R(t)=\int_{t}^{\infty}f(t)\mathrm{d}t=e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{4.25}$$

式（4.25）表明，对于起始时间为$t=0$的持续时间$t(=\gamma+\eta)$，不论β的值为多少，可靠度总是$R(t)=36.8\%$。因此，对于任何一个Weibull失效概率密度函数，产品存活$t(=\gamma+\eta)$时间段的概率都是36.8%。对于具有特定可靠度R的产品，其失效时间的表达式为

$$t=\gamma-\eta(\ln R)^{1/\beta} \tag{4.26}$$

Weibull分布的故障率函数为

$$h(t)=\frac{f(t)}{R(t)}=\frac{\beta}{\eta}\left[\frac{t-\gamma}{\eta}\right]^{\beta-1} \tag{4.27}$$

条件可靠度函数是

$$R(t,T)=\frac{R(t+T)}{R(T)}=\exp\left\{\left[\frac{(t+T-\gamma)^{\beta}}{\eta}\right]\left[\frac{(T-\gamma)^{\beta}}{\eta}\right]\right\} \tag{4.28}$$

公式（4.28）给出了产品完成持续时间为t的一个新任务的可靠性，T为新任务开始前已经累积的时间。由此可以看出，Weibull分布通常和任务开始时间以及任务持续时间（除非$\beta=1$）相关。实际上，除了指数分布以外，大部分分布都与任务开始时间及持续时间相关。表4.1列出了Weibull分布的关键参数。表中函数是伽马函数（Gamma Function），其值可以从统计表中获得。

表4.1　Weibull分布的关键参数

位置参数	γ
形状参数	β
比例参数	η
均值（算术平均值）	$\gamma+\eta\Gamma\left(\frac{1}{\beta}+1\right)$
中值（t_{50}或者50%失效的时间）	$\gamma+\eta(\ln 2)^{1/\beta}$
众数［$f(t)$的最大值］	$\gamma+\eta\left(1-\frac{1}{\beta}\right)^{1/\beta}$
标准偏差	$\eta\sqrt{\Gamma\left(\frac{2}{\beta}+1\right)-\Gamma\left(\frac{1}{\beta}+1\right)}$

Weibull分布的形状参数决定了故障率函数的形状。对于$0<\beta<1$，故障率是时间的函数，它代表早期寿命失效（也就是早期失效）；$\beta\approx1$表示故障率是一个常数，可以代

表“浴盆”曲线（Bathtub Curve）中的“使用寿命”阶段；$\beta>1$ 表示故障率上升，可以用它来表示磨损失效。如图 4.1 所示为 $\eta=1$ 和 $\gamma=0$ 时，β 对于概率密度函数曲线的影响；如图 4.2 所示为 $\eta=1$ 和 $\gamma=0$ 时，β 对故障率曲线的影响。

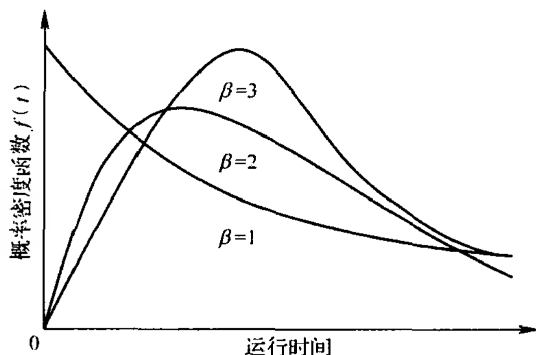


图 4.1　形状参数 β 对于概率密度函数的影响（$\eta=1$，$\gamma=0$）

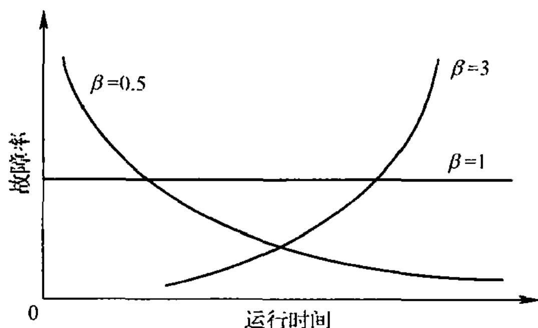


图 4.2　形状参数 β 对于故障率的影响（$\eta=1$，$\gamma=0$）

比例参数 η 起到放大或缩小时间轴比例的作用。因此，对于固定的 γ 和 β，η 的增大会使分布向右延伸，同时保持起始的位置和形状不变（另外，因为概率密度函数曲线下的总面积必须等于 1，所以振幅将会下降）。如图 4.3 所示为 $\beta=2$ 和 $\gamma=0$ 时参数 η 对于概率密度函数的影响。

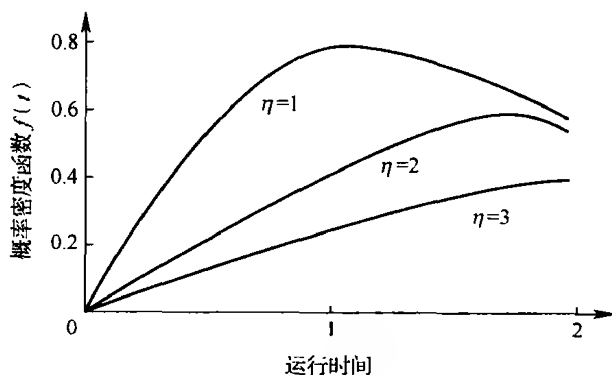


图 4.3　比例参数 η 对于 Weibull 分布的概率密度函数的影响（$\beta=2$，$\gamma=0$）

位置参数用来估计最早的失效时间，并定位其沿时间轴的分布。对于 $\gamma=0$，分布的起始时刻为 $t=0$；对于 $\gamma>0$，它表示产品在长度为 γ 的时间段内不会出现失效。如图 4.4 所示为 $\beta=2$ 和 $\eta=1$ 时 γ 对于概率密度函数曲线的影响。注意：如果 γ 是正数，分布从 $t=0$ 线左侧或者原点的右侧开始；如果 γ 是负数，分布从 $t=0$ 线左侧或者原点的左侧开始，并且表示失效在 $t=0$ 之前已经出现，例如在运输和存储阶段。Weibull 分布还可以用 $\gamma=0$ 的二参数分布表示。

Excel 中的 Weibull 分布函数：Weibull（x，β，η，Cumulative）返回二参数 Weibull 分布。

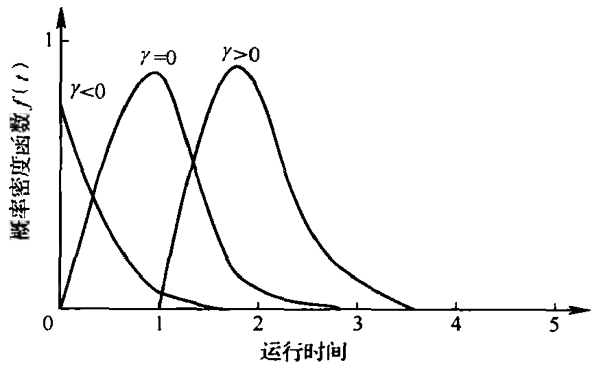


图 4.4　位置参数 γ 对于概率密度函数的影响
($\beta=2$，$\eta=1$)

x——用来估计函数的值，非负数；

β——分布函数的形状参数，正数；

η——分布函数的比例参数，正数；

Cumulative———一个逻辑值。对于累积分布函数，使用 TRUE。

4.3.2　指数分布

指数分布是一个单参数分布，简单易用。这种分布可以看做是 Weibull 分布的 $\beta=1$ 的特殊情况。指数分布用来对独立事件发生的恒定间隔时间进行建模。其概率密度函数为

$$f(t)=\lambda_0 e^{-\lambda_0 t}(t\geqslant 0) \tag{4.29}$$

式中，λ_0 是一个正实数，常称其为恒定失效率。表 4.2 总结了指数分布的关键参数。参数 λ_0 通常是未知的，并且必须对其进行计算或者估计。一旦 λ_0 已知，我们就可以用概率密度函数计算可靠度：

$$R(t)=\int_t^{\infty} f(\tau)\,\mathrm{d}\tau=\int_t^{\infty}\lambda_0 e^{-\lambda_0 t}\,\mathrm{d}\tau=e^{-\lambda_0 t} \tag{4.30}$$

累积分布函数或不可靠度的表达式为

$$Q(t)=1-\exp[-\lambda_0 t] \tag{4.31}$$

如上所述，故障率是一个常量：

$$h(t)=\frac{f(t)}{R(t)}=\frac{1}{e^{-\lambda_0 t}}(\lambda_0 e^{-\lambda_0 t})=\lambda_0 \tag{4.32}$$

表 4.2　指数分布参数

比 例 参 数	$1/\lambda_0$
中值（t_{50}或者 50% 失效的时间）	$0.693/\lambda_0$
众数 [$f(t)$ 的最大值]	0
标准偏差	$1/\lambda_0$
均值	$1/\lambda_0$

条件可靠度为

$$R(t,T)=\frac{R(t+T)}{R(T)}=\frac{e^{-\lambda_0(t+T)}}{e^{-\lambda_0 T}}=e^{-\lambda_0 t} \tag{4.33}$$

公式（4.31）表明，先前的试验（例如试验或者任务）对后来的可靠度没有影响。这种“和新的一样好”的结论来自于一个事实，即故障率是一个常数，且产品的失效概率和以往情况或者产品是否使用过无关。

服从指数分布的平均失效时间（Mean Time To Failure，MTTF）由连续型分布均值的一般方程计算：

$$\text{MTTF}=\int_0^\infty R(t)\,dt=\int_0^\infty e^{-\lambda_0 t}dt=\frac{1}{\lambda_0} \tag{4.34}$$

因为服从指数分布的故障率是一个常数，平均失效时间也就是平均失效间隔时间（Mean Time Between Failures，MTBF）。MTBF 和恒定失效率成反比，因此，可靠度可以表示为

$$R(t)=e^{-t/\text{MTBF}} \tag{4.35}$$

MTBF 有时候会被误解为产品的寿命或是 50% 产品失效的时间。对于 $t=\text{MTBF}$ 的任务时间，用公式（4.34）计算得到的可靠性 R（MTBF）$=0.368$。因此，只有 36.8% 的产品在任务中的存活时间等于 MTBF。

对于假定故障率是一个常数的可靠性试验，用累积的总任务时间除以相关失效总数，就可以得到 MTBF 的一个点估计值：

$$\text{MTBF}=t_a/r \tag{4.36}$$

式中，t_a 是总任务时间，r 是相关失效总数。

案例　4.6

证明指数分布是 Weibull 分布的一种特殊形式。

解：参考公式（4.22），令 $\beta=1$，$\gamma=0$，则

$$f(t)=\frac{1}{\eta}e^{-t/\eta}$$

因此，Weibull 分布简化为 $\lambda_0=1/\eta$ 的单参数指数分布。可靠度和故障率函数简化为

$$R(t)=e^{-t/\eta}$$

$$h(t)=\frac{1}{\eta}$$

式中，$\eta=\frac{1}{\lambda_0}$。

如果 $\beta=1$，$\gamma>0$，那么，Weibull 分布相似于延迟指数分布，我们可以把延迟看作不出现失效的周期。

案例 4.7

假设一个产品的失效时间可以用 Weibull 分布描述，其估计参数值 $\eta=1000\text{h}$，$\gamma=$

0，$\beta=2$。估计产品在工作100h后的可靠性，并确定MTTF。

解：由公式（4.21）和表4.1可得：

$$R(100)=e^{-(100/1000)^2}=0.990$$

并且，$MTTF=1000\Gamma(1/2+2)=1000\Gamma(1.5)=886h$。

案例 4.8

为下列可靠性试验估计MTBF：

（a）在没有替代品的情况下，采用失效截尾试验：对12个试件进行试验，直到出现第4次失效，失效分别出现在200h、500h、625h和800h。

（b）在没有替代品的情况，采用时间截尾试验：对12个试件进行试验，试验1000h，失效分别出现在200h、500h、625h和800h。

（c）在有替代品的情况下，采用失效截尾试验：对8个试件进行试验，直到出现第3次失效，失效分别出现在150h、400h和650h。

（d）在有替代品的情况下，采用时间截尾试验：对8个试件进行试验，试验1000h，失效分别出现在150h、400h和650h。

（e）有替代品和没有替代品的综合情况：对6个试件进行试验，试验1000h。第1次失效出现在300h，其替代品又工作400h后失效；第2次失效出现在350h，其替代品又工作500h后失效；第3次失效出现在600h，其替代品到试验结束也没出现失效。

解：

（a）$MTBF(e)=[(200+500+625+800+8(800))/4]h=2131h$；

（b）$MTBF(e)=[(200+500+625+800+8(1000))/4]h=2554h$；

（c）$MTBF(e)=[8(625)/3]h=1733h$；

（d）$MTBF(e)=[8(1000)/3]h=2667h$；

（e）$MTBF(e)=[(700+850+1000+3(1000)/5]h=1110h$。

案例 4.9

某电子产品只有一次“失效机会”，并且这种产品有恒定故障率。如果MTBF为5年，那么何时会有10%的产品出现失效？

解：使用公式（4.26），$R=0.90$，MTBF≈43800h（5年），求t即可，

那么，$t=-[(MTBF)\times\ln R]\approx4600h$，或者接近半年。

Excel中的指数分布函数：EXPONDIST（x，λ，Cumulative）

x——函数值，一个非负数。

λ——参数值，一个等于失效率常量的正数。

Cumulative——作为累积分布函数所返回函数的一个逻辑值（CDF或PDF/TRUE或FALSE）

4.3.3 正态分布

当随机变量受到一系列随机作用影响时，就会导致没有占主导地位的单个影响因素，此时随机变量就表现为正态分布。正态分布用来表示制成品的尺寸变化、材料性能和测量误差等，也可以用来评估产品的可靠度。正态分布的概率密度函数是

$$f(t)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\left(-\frac{1}{2}\right)\left(\frac{t-\mu}{\sigma}\right)^2\right](-\infty<t<+\infty) \tag{4.37}$$

其中，参数μ是均值或 MTTF，σ是分布的标准偏差。表 4.3 中列出了正态分布的参数，图 4.5 给出了概率密度函数的形状。

表 4.3　正态分布参数

均值（算术平均值）	μ
中值（t_{50}或者 50% 失效的时间）	μ
众数 [$f(t)$的最大值]	μ
位置参数	μ
形状参数	σ
s（σ 的估计值）	$t_{50}-t_{16}$

正态分布的累积密度函数或不可靠度为

$$Q(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^t\exp\left[\left(-\frac{1}{2}\right)\left(\frac{x-\mu}{\sigma}\right)^2\right]dx \tag{4.38}$$

我们称均值为 0、方差为 1 的正态随机变量为标准正态变量（z），其表达式为

$$z\equiv(x-\mu)/\sigma \tag{4.39}$$

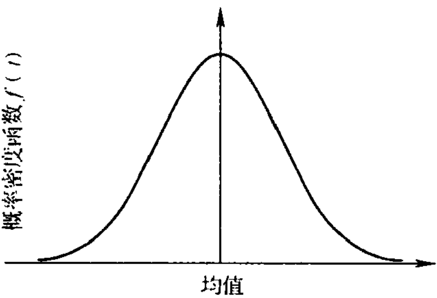


图 4.5　正态分布的概率密度函数

标准正态变量——特别是累积概率密度函数的特性——已在相关的统计表中列出。表 4.4 给出了由σ的等倍数确定的距离均值不同间距与标准正态曲线形成的面积占曲线下方总面积的百分比。

表 4.4　正态分布曲线下的面积

$\mu-1\sigma=15.87\%$	$\mu+1\sigma=84.13\%$
$\mu-2\sigma\approx2.28\%$	$\mu+2\sigma=97.72\%$
$\mu-3\sigma=0.135\%$	$\mu+3\sigma=99.865\%$
$\mu-4\sigma=0.003\%$	$\mu+4\sigma=99.997\%$

公式（4.34）的积分没有封闭解。因此，正态分布曲线下的面积可以从标准正态分布表中查得，用以下变换公式把变量转换为一个随机变量即可

$$F(t)=\Phi(z)=\Phi\left(\frac{t-\mu}{\sigma}\right) \tag{4.40}$$

标准正态分布用于描述总体存在预期磨损时间μ（通常定义为退化等级达到临界值的时间）的失效分布。轮胎表面的寿命和机床切削刃的寿命都属于类似情况。在此情况下，寿命由μ和标准偏差定义的不确定性——两者的平均值确定。在使用正态分布

时，失效出现在此平均时间前后的概率是相等的。

案例 4.10

某技师估计，洗衣机的空气压缩机在循环使用25000和35000次后，将有90%的可能会出现失效。假设洗衣机的老化服从正态分布，计算平均寿命和洗衣机寿命的标准偏差。

解：假定使用循环次数少于25000次的失效概率为5%、循环次数大于35000次的失效概率为5%，分布的均值将集中在循环30000次，那么，$\mu=30000$ 时：

$$\Phi(z_1)=0.05, z_1=\frac{25000-\mu}{\sigma}, \Phi(z_2)=0.95, z_2=\frac{35000-\mu}{\sigma}$$

从正态分布表中查得：$z_1=-1.65$，$z_2=1.65$。

因此，$-1.65\sigma=25000-\mu$，$1.65\sigma=35000-\mu$。

解这两个公式，得到标准偏差 σ 为3030次循环。

案例 4.11

腐蚀失效的出现时间服从正态分布，其均值为 $\mu=2.8\text{h}$，标准偏差为 $\sigma=0.6\text{h}$。

（a）1.5h内出现腐蚀的概率是多少？

（b）如果要分析10%增长量的腐蚀，需要在何时开始对真菌进行分析？

解：

（a）在1.5h之内，腐蚀增长概率的表达式为

$P\{t<1.5\}=Q(1.5)=\Phi(z)$

$z=(x-\mu)/\sigma=(1.5-2.8)/0.6=-2.1667$

从标准正态表中查得：$\Phi(-2.1667)=0.0151$。

（b）在此情况下，F（Φ）$=0.1$；那么，从标准正态表查得 z 的近似值为 -1.28。由于 $-t+\mu=0.28\sigma$，因此，$t=2.03\text{h}$。

Excel中的正态分布函数：

NORMDIST(x, Mean, Standard_ dev, Cumulative）——已定义均值和标准偏差返回正态累积分布。

NORMINV(Probability, Mean, Standard_ dev）——已定义均值和标准偏差返回正态累积分布的倒数。

NORMDIST(z)

NORMINV(Probability)

4.3.4 对数正态分布

对于一个连续型随机变量，它有可能是一系列随机变量的乘积。例如某系统的磨损可能与作用在系统上面的载荷量成一定比例，这种情况可以表示为

$$y=y_1y_2y_3\cdots y_N \tag{4.41}$$

式中，y_i 是不同的载荷，左侧的 y 代表磨损量。

对公式取自然对数：

$$\ln y=\ln y_1+\ln y_2+\ln y_3\cdots+\ln y_N \tag{4.42}$$

如果公式（4.39）右边没有对结果（也就是 $\ln y$ 的值）有决定性作用的单个项，那么 y 的分布呈正态，可认为 y 是对数正态分布。对数正态分布在很多工程情况下都有应用，如材料的强度、结构元素的尺寸、生理参数如骨骼关节的载荷等。在可靠性工程中，对数指数分布用来描述由疲劳引起的失效。其概率密度函数是

$$f(t)=\frac{1}{\sigma t\sqrt{2\pi}}\exp\left[\left(-\frac{1}{2}\right)\left(\frac{\ln t-\mu}{\sigma}\right)^2\right](0<t<\infty) \tag{4.43}$$

式中，σ 是所有失效时间对数值的标准偏差，μ 是所有失效时间对数值的均值。

对数正态分布的累积分布函数（不可靠度）是

$$Q(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^t\frac{1}{x}\exp\left[\left(-\frac{1}{2}\right)\left(\frac{\ln x-\mu}{\sigma}\right)^2\right]dx \tag{4.44}$$

σ 的两个值的概率密度函数如图 4.6 所示。表 4.5 中给出了对数正态分布的关键参数。

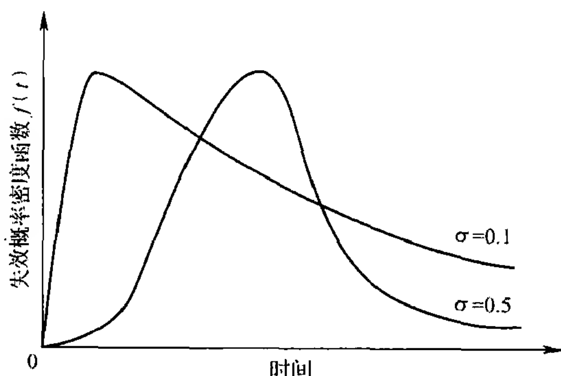


图 4.6　$\sigma=0.1$ 和 $\sigma=0.5$ 的对数正态分布的概率密度函数

表 4.5　对数正态分布参数

均值（算术平均值）	$T=\exp(\mu+\sigma^2/2)$
中值（t_{50}或者 50% 失效的时间）	$T=e\mu$
众数［$f(t)$ 的最大值］	$T=\exp[\mu-\sigma^2/2]$
位置参数	$e\mu$
形状参数	σ
s（σ 的估计值）	$\ln(t_{50}-t_{16})$

对于故障率服从对数正态分布的总体，其 MTTF 为

$$\mathrm{MTTF}=\exp\left(\mu+\frac{\sigma^2}{2}\right) \tag{4.45}$$

从对数运算符的基本性质可以看出，如果变量 x 和 y 服从对数正态分布，那么它们的乘积得到的随机变量 $z=xy$ 也服从对数正态分布。

案例 4.12

某工业断路器总体的失效服从参数 $\mu=3$、$\sigma=1.8$ 的对数正态分布。总体的 MTTF 是多少？持续工作 30 年后，这些断路器的可靠度如何？

解：由 MTTF 的公式（4.45）可得

$$\mathrm{MTTF}=\exp[3+0.5\times(1.8)^2]=101.5\ 年$$

对于持续使用 30 年可得

$$z=\frac{\ln(30)-3}{1.8}=\frac{3.41-3}{1.8}=0.228$$

因此，由标准正态分布表可知，持续工作 30 年后的可靠度为

$$R(30)=[1-\Phi(z)]=[1-\Phi(0.228)]=[1-0.589]=0.411$$

Excel 中的对数正态分布函数 LOGMORMDIST（x，Mean，Standard_ Dev，Cumulative）为已定义均值和标准偏差的 x 返回累积分布。其中，$\ln x$ 是正态分布，均值和标准差都是对 $\ln x$ 而言的。

4.4 绘制概率曲线

绘制概率曲线是一种确定数据（观测值）是否和假设分布一致的方法。我们通常用计算机软件来评估假设分布，并确定其概率参数。计算机软件工具所用的方法类似于用概率绘图坐标纸来为数据绘制曲线图。把失效时间数据按照合适的度量标准从小到大进行排列（例如失效时间、失效循环）；选择不可靠度的百分比估计值；把这些数据绘制到概率绘图坐标纸上（由分布类型指定的），x 轴为失效时间，y 轴为不可靠性百分比的估计值；画出通过这些数据点的最佳拟合直线。

x 轴上的失效时间来自于使用现场或试验。不可靠度的估计值与描绘这些失效时间数据的差异不是很明显，一些不同的技术例如“曲线中点绘图”、“绘制预期曲线位置”、“绘制曲线中值位置”和“Kaplan-Meier 分级”（软件中的）可在此类估计中使用。表 4.6 给出了根据不同估计方案为样本大小为 20 的总体估计不可靠度的方法。

表 4.6 $N=20$ 的 CDF 估计

	累积分布函数或不可靠度的估计值			
序 列	曲线中点位置	预期曲线位置	中值位置	中值等级
1	2.5	4.8	3.4	3.4
2	7.5	9.5	8.3	8.3
3	12.5	14.3	13.2	13.1
4	17.5	19.0	18.1	18.0
5	22.5	23.8	23.0	23.0
6	27.5	28.6	27.9	27.9
7	32.5	33.3	32.8	32.8
8	37.5	38.1	37.7	37.7
9	42.5	42.8	42.6	42.6

（续）

累积分布函数或不可靠度的估计值				
序　列	曲线中点位置	预期曲线位置	中 值 位 置	中 值 等 级
10	47.5	47.6	47.5	47.5
11	52.5	52.4	52.5	52.5
12	57.5	57.1	57.4	57.4
13	62.5	61.9	62.3	62.3
14	67.5	66.7	67.2	67.2
15	72.5	71.4	72.1	72.1
16	77.5	76.4	77.0	77.0
17	82.5	80.1	81.9	81.9
18	87.5	85.7	86.8	86.8
19	92.5	90.5	91.7	91.7
20	97.5	95.2	96.6	96.6

下面公式的解是中值等级：

$$\frac{N!}{i!\ (N-i)!}(1-Q)^{i}Q^{N-i}=0.5 \tag{4.46}$$

式中，N 是样本大小，i 是失效次数，Q 是中值等级（或者第 i 次失效所在时刻的不可靠度估计值）。公式（4.42）用来估计中值的绘图位置，它可以替代中值等级：

$$Q_i=\frac{100\times(i-0.3)}{N+0.4} \tag{4.47}$$

用于绘图的轴坐标不是线性的。每一个概率分布所用的坐标是不同的，并且由线性化的可靠性函数建立。通常，反复对公式的两边求对数可以实现函数的线性化。例如对于 Weibull 分布方程式（4.25）的数学处理，使得纵坐标（Y 轴）为倒数对数标尺，横坐标（X 轴）为失效时间的对数标尺。

一旦不同分布的概率图绘制完成，绘图点的拟合程度是确定哪个分布与数据最符合的依据。应该根据概率分布拟合数据的能力以及实际原因来选择数据分析的概率分布。分布是从引起失效的机理的失效模型中提取的。对于分布的选择，可能会存在一些实际意义方面的争议，选择分布类型的决策有时候很难制定。例如对数正态分布和 Weibull 分布都能很好地对疲劳失效数据进行建模。因此，这两者通常都可用来拟合失效数据，此时需要根据以往经验来作出工程决策。

认为所有的失效时间数据总和仅拟合一个失效分布是不合理的。因为引起产品失效的机理不止一种，不同的失效也可能是由一种机理引起的，而另外一些失效由另一种不同的机理引起。在这种情况下，没有概率分布能很好地拟合数据。即使某个分布能很好地拟合这些数据，它也不具备任何预测能力。这就是为何要根据机理对失效进行分类，然后为各个失效种类单独找出合适的分布。

表 4.7 给出了根据失效机理划分为两组的失效时间。图 4.7 显示了不同失效机理数据的 Weibull 概率曲线。需要注意的是：两组失效时间的曲线形状和比例因素是不同的，一组的故障率（$\beta=0.67$）逐渐下降，另外一组的故障率（$\beta=4.33$）逐渐上升。如果

把这些数据的曲线绘制在一起，结果显示的故障率几乎是恒定的。分析这两者相结合得到的数据，其结果可能会对制定备件和后勤决策引起误导，甚至会起到反作用。

表 4.7 由失效机理划分的失效时间

序号	状态 F 或 S	F 或 S 的时间	子组 ID
1	F	2	V
2	F	10	V
3	F	13	V
4	F	23	V
5	F	23	V
6	F	28	V
7	F	30	V
8	F	65	V
9	F	80	V
10	F	88	V
11	F	106	V
12	F	143	V
13	F	147	W
14	F	173	V
15	F	181	W
16	F	212	W
17	F	245	W
18	F	247	V
19	F	261	V
20	F	266	W
21	F	275	W
22	F	293	W
23	S	300	
24	S	300	
25	S	300	
26	S	300	
27	S	300	
28	S	300	
29	S	300	
30	S	300	
31	S	300	

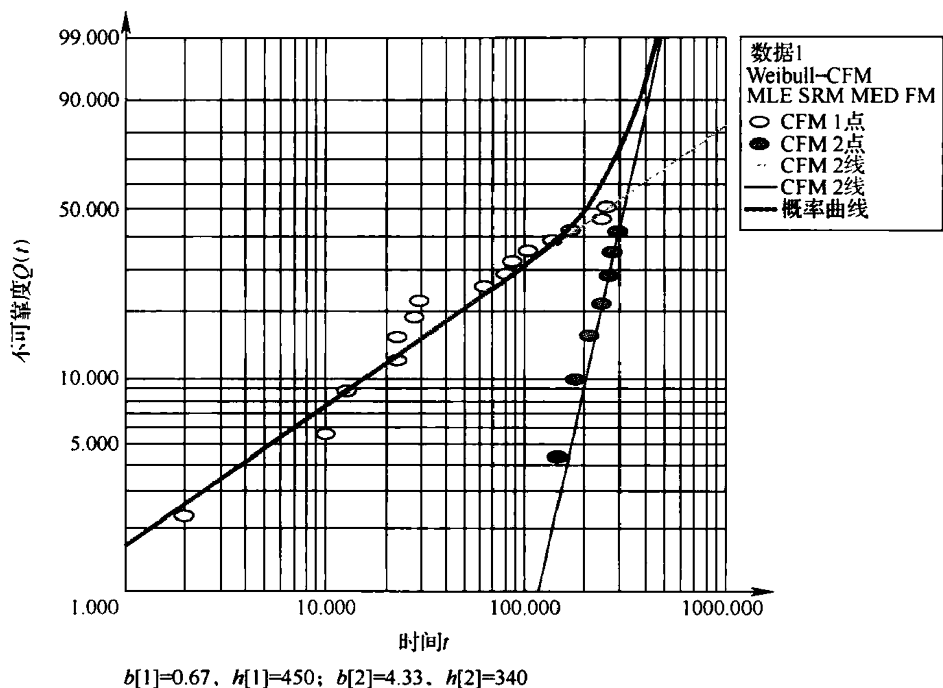


图 4.7　表 4.7 中失效机理数据的 Weibull 概率曲线

案例　4.13

表4.8给出了对10个相同产品的可靠性进行试验所得的数据，其中，6个产品在试验进行600h后失效。把失效寿命数据绘制在二参数Weibull分布绘图样上，用所绘制曲线（见图4.8）进行以下估计：

(a) 50h后的可靠度和不可靠度。

(b) 50h后，下一个50h周期的可靠度。

(c) 假定产品在50h后启动，仍有95%可靠性的最长周期。

表 4.8　案例 4.13 的试验数据

样本序号	失效时间/h	样本序号	失效时间/h
1	14	6	563
2	58	7	—
3	130	8	—
4	245	9	—
5	382	10	—

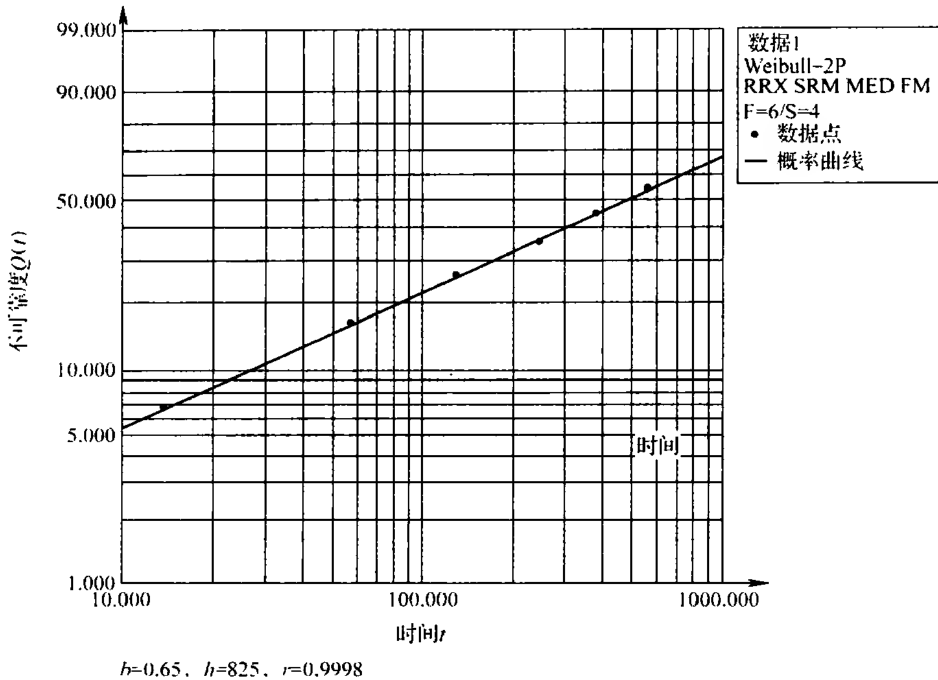


图4.8　表4.8中失效时间数据的二参数Weibull分布概率曲线

解：在此案例中，$\beta=0.65$，η 的估计值为825h。现在可以根据这些数据列出可靠度方程，用它来分析可靠度；也可以绘制直线来直接确定可靠度的值。

（a）从图4.7中可知，任务时间为50h的不可靠度估计值可以依据直线直接读出，即 $Q(50)=15\%$。因此，这个阶段的可靠度为 $R(50)=1-Q(50)=85\%$。

（b）50h后，下一个50h周期的可靠度可由条件可靠度方程给出，表达式为

$$R(50,50)=\frac{R(50+50)}{R(50)}=\frac{R(100)}{R(50)}=\frac{0.78}{0.85}=91.7\%$$

式中，$R(100)=1-Q(100)$ 可以从曲线直接得到。

（c）对于任务时间 t，它在50h周期之后仍有95%可靠度：

$$R(t,50)=\frac{R(t+50)}{R(50)}=\frac{R(t+50)}{0.85}=0.95$$

或者

$$R(t+50)=0.95\times0.85=0.808$$

为获得如此可靠度，产品的不可靠度为0.192或19.2%。从曲线中可以得出：此不可靠度值对应的时间大致为75h。那么，$50+t=75$，$t=25$，即为获得95%的可靠度，新任务的最长时间为25h。

寿命数据（Life Data）可能包括两个或更多寿命片段的数据，例如早期失效、有用寿命和磨损等，混合Weibull分布可用来拟合拥有不同分布参数的数据块。Weibull概率图样上的C线和S线（二参数或者三参数）可以表示混合Weibull分布的形状。

统计分析不能提供预测未来的神奇方法，分析的结果仅仅来自于假定模型以及假

设：例如如何定义失效、数据的有效性以及如何运用模型，都要考虑分布的尾部、推断和插值的局限性等。下面的例子显示了推断失效时间数据超出合理界限的现象。

案例　4.14

收集人口寿命（包括失效，即死亡）第一个 10 年的数据，将其作为一个总体，它的 Weibull 概率曲线图（如图 4.9 所示）。

（a）估计此总体在 300 年时失效的百分比。

（b）如果失效数据是人口死亡率，估计结果是否有意义？并进行讨论。

解：（a）图 4.9 的结果显示，总体在第 300 年时失效概率为 2%。

（b）百万人口在生命的第一个 10 年的死亡数据很好地拟合了 Weibull 分布，但此结果是错误的。从结果来看，很明显，这些数据不能用于对人类寿命进行任何判断，即便所有的计算都是正确的。在生命的第一个 10 年中，无法推断人类的死亡率，因为死亡率会随年龄变化（这对于工业产品也是正确的。在制造测试后出现的失效，通常是由制造缺陷引起的）。第一个 10 年的失效时间数据会导致形状参数（β）小于 1。但是，在幼年的早期到成年阶段的很大一部分时间，形状参数将非常接近 1，死亡可以认为是随机的（例如由事故引起）。然后，此总体将进入一个损耗阶段，引起死亡的原因是年龄。完整的人口死亡率数据应该使用混合 Weibull 分布来建模。

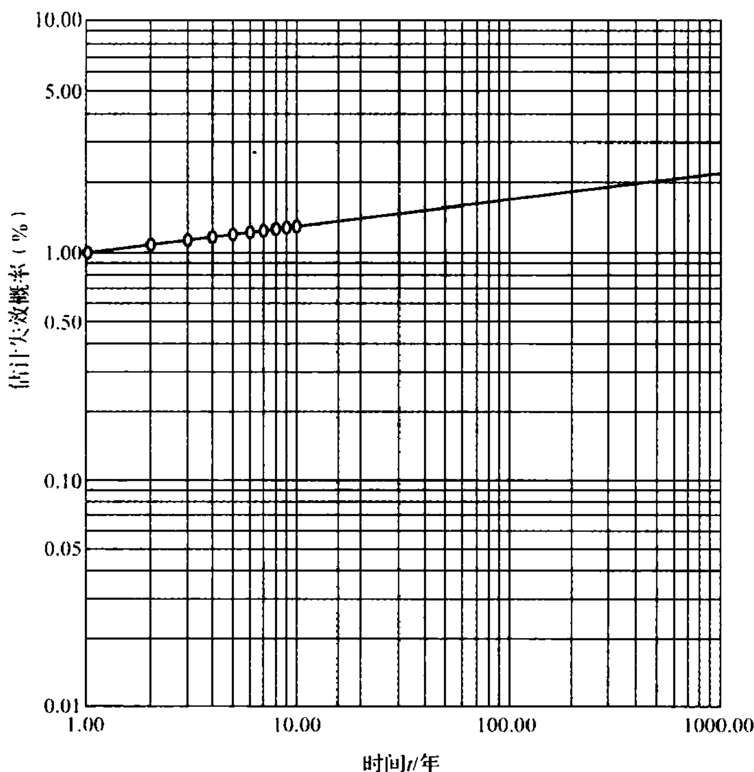


图 4.9　案例 4.14 失效时间数据的 Weibull 概率曲线图

第 5 章　置 信 区 间

5.1　引言

人们通常从总体的样本中收集一些数据，用以估计整个总体的某些特征。例如通过评估众多灯泡样本中一个样本的失效时间可以估计所有灯泡的寿命。另一个例子是对某个产品成本进行定期采样，就可以估计整个产品总体的缺陷率。也可以对产品进行样本验收测试，这样可以评估整个产品批次是否符合规格要求。

根据样本概括总体时会产生不确定性，置信区间（Confidence Interval）是度量此不确定性的一个量。本章将介绍置信区间的概念，并介绍在评估可靠性分析中利用置信区间评估不确定性的过程。

5.2　概念

总体（Population）是从某些数据群组的所有成员中收集的数据集合，样本（Sample）是仅从总体的一部分成员中收集的数据集合。从样本中获取的数据可以用来评估总体。通过评估一个样本来估计总体参数的过程如图 5.1 所示。此过程的先决条件是：总体必须是通过相同的过程创建的。度量整个总体是不可能的，我们也不建议这样做（例如某些度量行为可能会有损于某个样本）。从一个样本中计算而来的参数称为参数的点估计（Point Estimate）。置信区间为这些点估计设定一个界限，并提供了总体参数包括在这些界限内的概率。

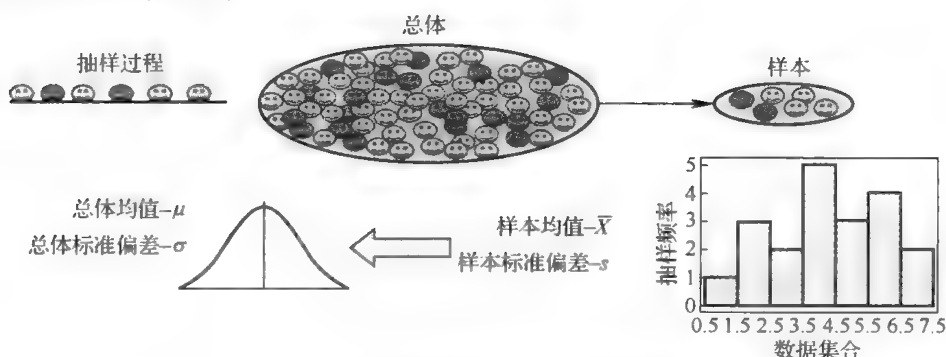


图 5.1　从样本参数估计总体参数的过程

推论统计（Inferential Statistic）用来从一个样本中为总体提出推论。对一个样本的统计，包括样本的位置，如平均数、中数和众数等；样本的波动，如方差、标准差、极差和四分位差等。

置信区间（Confidence Interval）是一个范围，它是从已知具有一定确定性实际值的样本参数计算而来的。置信区间的宽度表示了实际参数的不确定性。

计算被测数据集合的标准偏差（Standard Deviation）与估计其置信区间是不同的。标准偏差是一个度量数据分散程度的指标。一般来讲，标准偏差越高，度量数据均值的置信区间就越宽。然而，对于一个数据集合的统计不仅仅包括标准偏差。在实际中，虽然为参数分布计算标准偏差有可能是毫无意义的，但为其估计置信区间却具有一定的实际意义。例如对于一个数据集合，我们只可能计算其标准偏差的一个值，而不可能通过估计值得到标准偏差。

5.2.1　定义

当 θ 在 l 和 u 中间的概率为 $P(l \leqslant \theta \leqslant u) = 1 - \alpha (0 \leqslant \alpha \leqslant 1)$ 时，区间 $l \leqslant \theta \leqslant u$ 称为 $100 \times (1-\alpha)\%$ 置信区间。在此定义中，l 为下置信限；u 为上置信限；$1-\alpha$ 称为置信水平，通常以百分数形式表示。

一个置信区间要么是单边的，要么是双边的。一个双边（或双尾）置信区间为参数的区间估计定义了一个上限和下限，一个单边（或单尾）置信区间仅为参数的区间估计定义了一个下限或上限。一个只有下限、水平为 $100 \times (1-\alpha)\%$ 的单边置信区间为 $l \leqslant \theta \leqslant u$，其中，$l$ 的选择范围为 $P(l \leqslant \theta) = 1 - \alpha$；同样，一个只有上限、水平为 $100 \times (1-\alpha)\%$ 的单边置信区间为 $\theta \leqslant u$，u 的选择范围为 $P(\theta \leqslant u) = 1 - \alpha$。

5.2.2　置信水平

通常的观点是：置信水平是一个参数落在置信区间内的概率。虽然这种假设具有一定的主观性，并且有助于理解，但置信区间的概念不仅仅局限于此。一本工程统计教科书［Montgomery and Runger，1994］把置信区间详细描述为：实际中，我们仅能得到一个随机样本，只能计算一个置信区间。无论此区间是否包括 θ 的实际值，特定事件指定一个概率水平都是不合理的。适当的表述应该是：观测区间［l，u］包括 θ 的实际值，它的置信水平为 $100 \times (1-\alpha)$。此表述会受到抽样频率的影响，也就是说，对于特定样本，我们不能确定以上表述是否为真，但用于获取区间［l，u］的方法，有 $100 \times (1-\alpha)\%$ 的概率能产生正确的结果。

从一个置信水平为95%的总体中提取一些样本，从这些样本中计算的均值的50个置信区间如图5.2所示。其中，实线表示从整个样本总体中计算的实际均值。我们预计所有从总体中提取样本的95%将产生一个包括所估计参数的真实值的置信区间，只有5%的样本产生的置信区间不包括所估计参数的真实值。此模拟案例表明，置信区间的

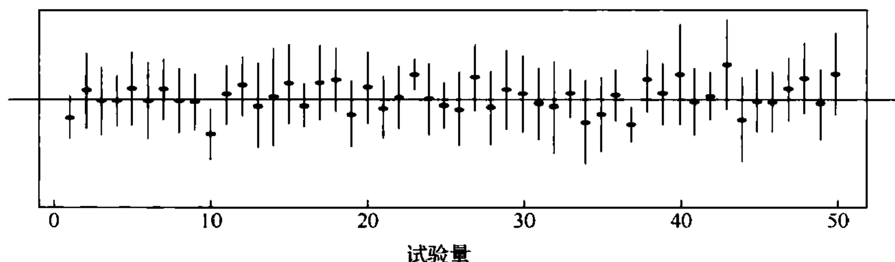


图5.2　置信区间的概念化表示

3个部分（大约5%）不包括参数的真实均值。

当样本量一定时，置信水平越高，置信区间就越宽。一个置信水平为100%的置信区间一定包括未知参数的真实值；然而，置信区间将放大到（$-\infty$，$+\infty$），这么大的置信区间不能提供任何有用信息。例如你可以以非常高的置信水平说：所有上可靠性课的学生的年龄都在1～100之间，但这不能为决策的制定提供任何有用信息。

置信水平的选择是工程风险分析过程的一部分。例如通过置信区间分析，工程人员可以估计一段时间内所能预期的最坏的返修情况。通过对预期返修的置信水平的95%或99%（或其他由工程人员选定的任意值）进行点估计，相关人员能计算所要储备的备件数量。根据备件的成本与由于备件不可用造成的维修时间延迟成本之比，相关人员就可以制定决策了。在很多情况下，行业惯例或客户合同可能会要求特定的置信水平，常见值为90%或95%。

5.2.3　置信区间和样本量的关系

置信区间的值取决于每个样本的大小。假定置信水平保持恒定（不变），只要从相同的总体中提取样本量，增加样本量就会使置信区间的宽度变小。然而，当执行某种测试或从现场收集数据时，数据可能会来自多个总体。在此情况下，大样本量或许会拓宽置信区间。例如在棒球棒制造商中，工程人员可能会记录从生产线上提取样本的硬度值。如果并不是所有的生产参数都受控，那么所有来自相同总体的样本量的增加将使置信区间变窄；但是，如果在特定时间段内生产参数不受控，在此期间内提取样本的硬度值将会有所不同。因此，增加来自“不受控”总体的样本量将使置信区间变宽。

5.3　置信区间估计方法

常用于估计置信区间的计算方法有两种：Fisher矩阵（Fisher Matrix，FM）界限法常用于大部分商业统计中，因为其结果太过乐观，所以不推荐在小样本量中使用；似然比（Likelihood Ratio，LR）置信界限法可用于任何样本量，然而在样本量非常大时，其计算强度会让LR分析变得非常耗时。

以上两种方法的数学表达式及其变形超出了本书的讨论范围。用Fisher矩阵方法确定的Weibull分布的形状参数和尺度参数的置信界限为

$$
\begin{aligned}
\beta_{U} &= \hat{\beta} \times e^{\frac{K_{\alpha}\sqrt{Var(\hat{\beta})}}{\beta}} \\
\beta_{L} &= \frac{\hat{\beta}}{e^{\frac{K_{\alpha}\sqrt{Var(\hat{\beta})}}{\beta}}} \\
\eta_{U} &= \hat{\eta} \times e^{\frac{K_{\alpha}\sqrt{Var(\hat{\eta})}}{\eta}} \\
\eta_{L} &= \frac{\hat{\eta}}{e^{\frac{K_{\alpha}\sqrt{Var(\hat{\eta})}}{\eta}}}
\end{aligned}
\tag{5.1}
$$

其中，K_α 计算为

$$\alpha = \frac{1}{2\pi}\int_{K_\alpha}^{\infty} e^{\frac{t^2}{2}} dt$$

似然比法的计算公式为

$$L(\beta,t) = \prod_{i=1}^{N} \frac{\beta}{\left(\frac{t}{(-\ln R)^{\frac{1}{\beta}}}\right)}\left(\frac{x_i}{\left(\frac{t}{(-\ln R)^{\frac{1}{\beta}}}\right)}\right)^{\beta-1} \exp\left[-\left(\frac{x_i}{\frac{t}{(-\ln R)^{\frac{1}{\beta}}}}\right)^{\beta}\right] \tag{5.2}$$

5.4 正态分布的置信区间

置信区间的相关概念通常用正态分布来描述，其中一个原因是，置信区间是一个由两参数描述的对称分布。在服从正态分布的总体中，置信区间与样本量大小有直接的关系。

本节将介绍三种情况下的置信区间的计算：已知方差，未知均值总体的置信区间；未知方差，未知均值总体的置信区间；已知方差，但两总体均值不同时的置信区间。

5.4.1 已知方差，未知均值总体的置信区间

假定某总体均值μ未知，但方差σ^2已知。方差可从历史数据中获得，如创建总体或控制图的实际过程。对于此总体，设大小为n的随机样本产生的样本均值为X。总体均值$100\times(1-\alpha)\%$的置信区间可计算为

$$X-\frac{Z_{\alpha/2}\sigma}{\sqrt{n}}\leqslant\mu\leqslant X+\frac{Z_{\alpha/2}\sigma}{\sqrt{n}} \tag{5.3}$$

其中，$Z_{\alpha/2}$是标准正态分布曲线上方（$\alpha/2$）%的点。为了获得单边置信区间，要相应地用Z_α代替$Z_{\alpha/2}$。分别设定$l=-\infty$，$u=+\infty$，单边置信区间可计算为

$$\mu\leqslant u = X+\frac{Z_\alpha\sigma}{\sqrt{n}} \tag{5.4}$$

和

$$X-\frac{Z_\alpha\sigma}{\sqrt{n}} = l\leqslant u \tag{5.5}$$

当使用样本均值X估计实际未知均值μ时，误差为$E=|X-\mu|$，此误差不会立即显现出来。当双边置信区间的水平为$100\times(1-\alpha)\%$时，误差可以由以下公式精确计算：

$$E\leqslant\frac{Z_{\alpha/2}\sigma}{\sqrt{n}} \tag{5.6}$$

因此，我们可以选择大小为n，使得置信水平为$100\times(1-\alpha)\%$的一个样本量，误差将不会超过规定量E：

$$n=\left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2 \tag{5.7}$$

其中，n 舍入到下一个整数。

案例 5.1

假设要度量一个数字电子零件的传输延迟。你想要的置信水平为99%，即所得的传输延迟均值与实际传输延迟均值的误差在0.15ns内，应该如何选择样本量？

解：使用公式（5.9），并把 n 的值设定为34，则

$$n=\left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2=\left(\frac{Z_{0.995}\times 0.35}{0.15}\right)^2=\left(\frac{2.51\times 0.35}{0.15}\right)^2\approx 34$$

在此问题中，α 为0.01，$\alpha/2$ 则为0.005。因为标准正态分布是对称的，所以，$Z_{0.005}$就等同于$Z_{0.995}$。相应值可以从标准正态分布表中查得。

5.4.2 未知方差，未知均值总体的置信区间

学生 t 分布（Student's t-distribution）是一种用以估计分布展开的方法。此分布用来估计一个不服从正态分布假设的样本。如果总体被“假设”为正态分布，t 分布可作为均值的样本分布使用，样本方差 s^2 可用来取代未知的总体方差 σ^2。

假设某总体拥有未知方差 σ^2，大小为 n 的随机样本产生的均值为 X、样本方差为 s^2，t 分布曲线上方（$\alpha/2$)%的点拥有的自由度为（$n-1$）。此时，置信水平为 $100\times(1-\alpha)\%$ 的双边置信区间可计算为

$$X-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\leqslant\mu\leqslant X+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}} \tag{5.8}$$

案例 5.2

对用以制造安全带的合成纤维来说，在预测安全带的可靠性时，其拉伸强度是一个重要的特征值。根据以往经验，我们可以假设拉伸强度服从正态分布。从一批纤维中随机抽取16个样本进行测试。经计算，样本拉伸强度的均值为49.86psi，其标准偏差为1.66psi。为估计整批纤维的拉伸强度均值决定一个合适的区间。

解：因为工程人员可能只关注于那些非常低的拉伸强度值，那么为整批纤维的均值 m 选择单边置信区间是比较合适的。总体（整批纤维）的方差未知，且样本量相当小，需要根据 t 分布建立一个置信区间。总体均值 μ 的置信水平为99%的单边置信区间为

$$P\left\{-t_{\alpha,n-1}\leqslant\frac{\bar{x}-\mu}{\sqrt{s^2/n}}\right\}=1-\alpha\Rightarrow P\left\{-t_{0.01,1.5}\leqslant\frac{49.86-\mu}{\sqrt{1.66^2/16}}\right\}=0.99$$

$$\Rightarrow 49.86-\frac{(1.753)1.66}{\sqrt{16}}\leqslant\mu\Rightarrow 49.13\leqslant\mu$$

5.4.3 已知方差，但均值不同的两个总体的置信区间

两正态分布均值差的置信区间定义了一个范围，此范围内的值可能包括两总体均值的差（$\mu_1-\mu_2$）。从第一个总体中随机抽取一个样本 n_1，其已知标准差为 σ_1，产生的样本均值为 X_1；同样，从第二个总体中随机抽取一个样本 n_2，其已知标准差为 σ_2，产生

的样本均值为 X_2。那么，这两个样本均值差的水平为 100 × (1 - α)% 的双边置信区间为

$$X_1 - X_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leqslant (\mu_1 - \mu_2) \leqslant X_1 - X_2 + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \tag{5.9}$$

其中，$Z_{\alpha/2}$是标准正态分布上方（$\alpha/2$）% 的点。

案例　5.3

测试两种不同类型的用于焊接电子设备电源的铝线的拉伸强度。测试结果如表 5.1 所示。两种铝线的强度均值差（$\mu_1 - \mu_2$）的置信水平为 90% 的区间范围如何？

表 5.1　两种铝线的测试结果

类　型	样本量/n_i	样本拉伸强度均值/(kg/mm²)	总体的已知标准偏差/(kg/mm²)
1	15	86.5	1.1
2	18	79.6	1.4

解：从已知条件，可得

$$l = X_1 - X_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$= \left[86.5 - 79.6 - 1.645\sqrt{\frac{(1.1)^2}{15}+\frac{(1.4)^2}{18}}\right]\text{kg/mm}^2 = (6.9 - 0.716)\text{kg/mm}^2$$

$$= 6.184\text{kg/mm}^2$$

所以

$$u = X_1 - X_2 + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$= \left[86.5 - 79.6 + 1.645\sqrt{\frac{(1.1)^2}{15}+\frac{(1.4)^2}{18}}\right]\text{kg/mm}^2 = (6.9 + 0.716)\text{kg/mm}^2$$

$$= 7.616\text{kg/mm}^2$$

5.5　MTBF 的置信区间——假设为指数分布

假设平均故障间隔时间（Mean Time Between Failures，MTBF）服从指数分布，其置信界限可计算为

$$\text{MTBF} = \frac{2T_a}{\chi^2_{\gamma;dF}} \tag{5.10}$$

其中，T_a 是案例 3.6（参见第 3 章）中计算得出的总装置时（Device-Hours），χ^2 分布的参数 γ 和 dF（Degrees Of Freedom，自由度）的值可从表 5.2 中根据不同的试验条件查得。

表 5.2 用以计算 MTBF 置信界限的参数 γ 和 dF 的值[①]

试验类型	MTBF (l)		MTBF (u)	
	γ	dF	γ	dF
双边失效截尾试验	$\alpha/2$	$2r$	$1-\alpha/2$	$2r$
单边失效截尾试验	α	$2r$	$1-\alpha$	$2r$
双边时间截尾试验	$\alpha/2$	$2r+2$	$1-\alpha/2$	$2r$
单边时间截尾试验	α	$2r+2$	$1-\alpha$	$2r$
未观测到失效	α	2	—	—

① γ 为观测失效次数。

被测对象的失效服从指数分布，它的可靠度与所有等长间隔的内的可靠度相同，无论试验何时开始，可靠度都可通过公式（5.11）进行计算：

$$R = e^{\frac{-t}{\text{MTBF}}} \tag{5.11}$$

案例 5.4

在失效截尾试验（Failure-Terminated Test）中一共有 4 次失效，累积装置时为 16000 小时。

（a）MTBF 水平为 90% 的单边置信界限的上、下限分别是什么？

（b）100h 内，可靠性置信水平为 90% 的单边置信界限是什么？

解：由于

$T_a=16000\text{h}$，$CL=1-\alpha=0.90$，$\alpha=0.10$，$\alpha/2=0.05$，$1-\alpha/2=0.95$，$r=4$，

因此

$$\text{MTBF}(l)=\frac{2(16000)}{\chi^2_{0.10;8}}=\frac{32000}{13.362}=2395\text{h}$$

$$\text{MTBF}(u)=\frac{2(16000)}{\chi^2_{0.90;8}}=\frac{32000}{3.490}=9195\text{h}$$

如果被测对象的 MTBF 的置信界限为 0.90，其上、下限分别为 9195h 和 2395h。在所有长度为 100h 的间隔内，可靠性的上、下界限为

$$R(l)=e^{\frac{-100}{2395}}=e^{-0.0417}=0.9591$$

$$R(u)=e^{\frac{-100}{9195}}=e^{-0.0109}=0.9891$$

案例 5.5

在时间截尾试验（Time-Terminated Test）中有 7 次失效，累积装置时为 21000h。置信水平为 0.99 的 MTBF 的单边上、下限如何？

解：因为

$T_a=21000\text{h}$，$CL=1-\alpha=0.99$，$\alpha=0.01$，$\alpha/2=0.005$，$1-\alpha/2=0.095$，$r=7$，

所以

$$\text{MTBF}(l)=\frac{2(21000)}{\chi^2_{0.01;16}}=\frac{42000}{32000}\text{h}=1313\text{h}$$

$$\mathrm{MTBF}(u)=\frac{2(21000)}{\chi^2_{0.99;14}}=\frac{42000}{4.660}\mathrm{h}=9013\mathrm{h}$$

由于试验类型是时间截尾试验，所以无法建立置信上限，因为失效可能会立即发生，或是刚好在下一个测量区间的开始处发生。

常见的情况是，对 MTBF 以及相关的置信区间进行估计时，没有失效发生。但我们仍然能够计算单边置信界限的下限，它是 MTBF 的一个保守值。此处的测试是时间截尾试验，它假定失效会刚好在下一刻发生，或在下一个测量区间的开始处发生。当然，没有公式能计算单个（假想）失效的单边置信界限的下限，它也没有单边置信上限。

5.6　总体比例置信区间

在工程应用中，产品的检出质量并不是根据零件样本的测试进行估计的。如果 $\hat{p}$ 是大小为 n 的随机样本的观测比例，样本大小 n 根据某一种测试目的（例如缺陷）确定，那么，对于属于此类总体的比例为 p 的样本，其水平近似于 $100\times(1-\alpha)\%$ 的置信区间计算为

$$\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\leqslant p\leqslant\hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\tag{5.12}$$

其中，$Z_{\alpha/2}$是标准正态分布上方（$\alpha/2$）% 的点。当观测比例不是十分接近 0、1 或者样本量 n 时，这种关系是成立的。

案例　5.6

某检查员从生产线上随机选择 200 个配电盘，发现其中 5 个有缺陷。为从生产线上提取完好配电盘的比例计算水平为 90% 的置信区间。

$$\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\leqslant p\leqslant\hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{195}{200}-1.64\sqrt{\frac{0.975(0.025)}{200}}\leqslant \mathrm{p}\leqslant\frac{195}{200}+1.64\sqrt{\frac{0.975(0.025)}{200}}$$

$$0.957\leqslant p\leqslant 0.993$$

结果表明，总体样本中好的配电盘的比例在 0.997 和 0.993 之间。注意：我们没有为总体样本做出任何假设。

5.7　总结

用确定程度（或不确定程度）来描述所有工程分析的结果是一种比较好的方法，置信区间就是这样的一个统计量。本章描述了与置信区间相关的概念。在实际的工程应用中，这些概念可用来为可靠性分析和报告中的计算量估计置信区间。例如它们可以为回归参数计算置信区间。正如置信区间可以用来估计一些未知参数的分布一样，回归参数的置信区间也可以用来估计回归关系的不确定性。统计软件工具可以为计算量或由图

表生成的量自动计算置信区间。此类软件让报告更加容易，且使用者不需要完全理解这些值。

当把置信区间当作一个不确定量时，完整地报告相关信息是必不可少的。无论区间是单边还是双边的，报告都一定要包括置信水平。通常，报告最好包括样本量以及样本的选择过程。在正式的工程文献中，还需要提及所使用的分析方法（例如 Fisher 矩阵）。置信区间并不是唯一表达数据不确定性的方法。在一些情况下，我们可能无法估计并可视化置信区间。例如样本量非常小，工程人员就可能会获得一个没有实际使用价值的非常宽的置信区间。此时，数据可视化技术可以显示完整的结果，且不需要任何统计方法，它对相关人员根据数据制定决策有一定的帮助。

参考文献

Montgomery, D., and G. Runger. 1994. Applied statistics and probability for engineers, 324. New York: John Wiley & Sons.

第 6 章　硬件可靠性

6.1　引言

可靠性评估及其相关的验证技术是所有工程硬件成功运行的关键所在。通常，硬件可靠性的定义为：在整个预期任务寿命周期中，产品在特定生命周期载荷下，特定的容限内运行的概率（此处包括所有影响硬件性能的外部因素，例如机械力、温度、时间、有害化学物质、放射性物质和电压/电流等）。可靠性评估的目的是为选择影响可靠性（和受可靠性影响）的行为提供标准。可靠性不是偶然形成的，必须通过仔细定义良好的设计和制造措施，要有意识地、主动地把它内建在硬件产品中。在设计阶段，先量化地进行可靠性评估是实现其他各种设计功能的有效工具。这些设计功能包括：

① 基于复杂性、成本和风险的可靠性定位。

② 可行性评估。

③ 基于材料特性、应用剖面数据或现场失效数据等相关数据确定缺陷。

④ 基于相对可靠性裕度对可选设计进行配置。

⑤ 基于相对可靠性裕度对可选制造工艺进行比较。

⑥ 基于相对可靠性裕度对成本效益进行评估。

⑦ 产品参数，如成本、风险、开发时间、可生产性和维修性之间的权衡。

⑧ 为了让产品质量达到客户需求设计加速试验。

⑨ 为工艺验证设计加速应力试验。

⑩ 为修复行为确定可靠性问题。

⑪ 基于成本和风险的权衡，对降额和冗余设计进行决策。

⑫ 后勤计划的制订，如维修性决策。

⑬ 通过监测可靠性增长度量产品开发进度。

⑭ 维修保障分析。

各种因素之间错综复杂的关系是导致硬件失效的原因，这些因素包括应力（即作用在模块上及其内部的环境参数）和零部件的材料、配置、连接方式、装配关系等。要对可靠性进行适当的评估，就要系统地分析材料和产品配置对应力的反应。从失效物理来讲，产品失效的特征是由失效模式、位置和机理所定义的。

失效模式是一种观察到的物理变化，它由断路、短路、阻值增加或电子产品中其他电学参数改变之类的失效机理引起。在零部件质量检测中，可以把失效模式作为一个评判标准。传统的失效模式影响分析（Failure Mode Effect Analysis，FMEA）是检测失效表

现并对其评级的有效工具。失效机理是热、机械、电学、化学和磁应力相结合的过程，它会引发产品失效。失效过程通常从现有缺陷开始，如材料孔洞或两材料接触面间的微裂纹。一般来说，失效机理会出现在任意位置。虽然失效机理、失效模式和初始缺陷都被称为失效根源，但不能将它们混为一谈。

正确的可靠性评估程序是：首先对所有关键失效机理进行调查，然后确定这些机理出现的时间和位置，并确定这些机理在所要求的设计寿命内对产品运行产生的影响。由于失效机理可能出现于很多位置，在确定失效时，必须考虑运行时间内可能被激活的潜在失效机理。在零部件的质量检测过程中，我们必须要了解有可能出现的失效机理以及导致机理出现的应力，这样才能保证部件或组件能够承受这些应力。因此，对失效机理的调查研究也可以作为零部件质量检测的指导。

有效的量化模型不仅可以模拟潜在和相关的失效机理，还可以适应统计技术模型中的参数变化，它可以让可靠性评估变得更加简单。此类模型涉及在已知载荷情况下调查失效位置所承受的应力。应力是失效位置的载荷强度，基于失效位置处的几何配置和材料构成特征，可以使用合适的应力分析对它进行计算，例如基于相关材料的导电和导热性能，对某电子箱进行的应力分析可能包括计算特定降温条件下，设备门周围区域的温度；然后就可以基于应力等级，利用一个合适的通用材料损伤模型对失效进行预测。这就是“失效物理方法”。通常，输入到一个失效模型的数据是局部应力和材料的适当失效特征。如果失效机理为疲劳，那么就需要材料的疲劳特征和周期性机械应力；如果失效机理为电迁移，当应力是当前电流密度和局部温度的结合应力时，所需材料特征就包括激活能和速率常数。

不适当的过程控制和工艺会引发材料的变异性。在失效位置处的材料性质、几何公差、载荷以及预先存在的缺陷中，天然材料的变异性和瑕疵是最常见的现象。材料中段与段之间的变化很难用确定的模型来表达，通常较方便的做法是在可靠性建模中引入概率机制方法。在此方法中，每个参数的变异性都由合适的随机分布函数表达。然而，这只是一种表面特征（Phenomenological Feature)，数学分布的精确程度取决于所实施的观测量。统计分布尾端可达到的精确程度限制了可靠性评估的准确性。

本章将分节讨论以下内容：

① 工程硬件中常见的失效机理种类。

② 在每个失效位置处进行应力级别评估的分析技术。

③ 可靠性的质量检测和检验技术。

④ 用于探究制造工艺对质量和可靠性影响的质量评估技术。

⑤ 基于失效机理模型的硬件可靠性预测过程案例。

我们对失效模型和案例进行了尽可能详细和明确的阐述。然而，在可靠性评估过程中有很多不确定因素，对现场失效数据进行全面验证是必不可少的过程。因此，在新（新兴）技术中，可靠性评估模型通常会被迭代多次，以便基于实际失效经历对其进行不断的改进。

6.2　失效机理和损伤模型

在已知生命周期载荷历史的情况下，量化失效评估包括对预期载荷进行应力分析；失效机理建模；基于所有输入参数的随机变异性，对参数化和敏感度进行分析；把分析结果表示成有效的时间关联概率（Time-Dependent Probability）。要建立精确的可靠性模型，就要对潜在的应力和失效机理有清晰的认识。可靠性模型的作用效果取决于以下各因素的精度：

① 材料失效模型。

② 机械应力分析。

③ 用于分析的经验性数据（失效位置的几何状况和材料特性数据库）。

失效机理是应力损伤产品材料的物理过程。研究失效机理有利于无故障、可靠性设计的开展。如果量化模型能用于描述相关的失效机理，那么建立可靠性评估模型的过程将变得更加容易。因此，首先要确定那些在典型的工程硬件中，在生命周期内出现的应力触发的失效机理。

表 6.1 列出了常见的失效机理类型，它们都可能引发产品失效。关于这些失效机理的详细资料，可以在参考文献［Dasgupta A and J M Hu，1992］中找到。这些失效机理都是精心挑选出来并根据损伤特性进行描述的，以保证每一种都有一个通用的量化失效模型，常见工程材料的相关文献对它们进行了合理的描述。表 6.1 中的失效机理根据引发失效的类型分成两组。由于单独出现的应力超过材料内在强度而引发的灾难性突发失效称为过应力失效机理；当累积损伤超过材料的耐久度，就会出现增量损伤，由于增量损伤的单调累积引发的失效称为磨损失效机理（Wearout Mechanism）。

表 6.1　常见失效机理

任何单一应力偏移超过强度时的过应力失效	累积损伤超过耐久度时的磨损失效
1）性能失效与材料损伤无关	材料失效机理
① 机械失效	① 疲劳
② 电失效	② 蠕变
③ 热失效	③ 金属迁移
④ 表面失效	④ 腐蚀
2）材料失效机理	⑤ 磨损
① 断裂	⑥ 老化
② 扭曲	a. 相互扩散
③ 屈服	b. 解聚
④ 界面断裂	c. 脆化
⑤ 电气过应力	
⑥ 静电放电	
⑦ 介电击穿	
⑧ 热击穿	

意料之外的大应力事件可导致灾难性的过应力失效，也会使磨损损伤进一步累积，从而缩短产品寿命。这类应力的例子包括偶然的违规操作和天灾等。另外，设计精良、高质量硬件中的预期设计应力仅会引起磨损损伤的均匀累积，引发最终失效所需的损伤阈值不应该出现在设计寿命中。磨损失效机理的例子包括电子产品的功率循环中由热机械应力引发的疲劳损伤、由预期污染物引发的腐蚀速率、高功率设备中的电迁移速率等。

引发表6.1中失效机理的应力包括化学、热、电/磁、辐射和机械应力。引发机械失效的原因包括弹性和塑性形变、扭曲、脆性和（或）塑性断裂、界面分离、疲劳裂纹的发生和扩散、蠕变和蠕变断裂。引发热应力过载失效的原因是零件温度超过了临界温度，例如玻璃化温度、熔点、虚化点或燃点。热磨损失效的例子包括解聚（线型高聚物降解为低聚物或单体的过程）引起的老化、金属间的增长和相互扩散。电气失效通常出现在电子硬件中，包括由电气过应力（Electrical Overstress，EOS）和静电放电（Electrostatic Discharge，ESD）引发的应力失效和磨损失效，如介电击穿（Dielectric Breakdown）、结击穿（Junction Breakdown）、热电子注入、表面和体积凹陷、表面击穿等。电迁移就是一种磨损失效。

辐射失效主要由铀和钍的污染物以及次级宇宙射线引起，它会引起磨损、老化、材料脆化和电子硬件（如逻辑芯片）的过应力“软”错误。化学失效发生在恶劣的化学环境中，它会导致腐蚀、氧化或离子表面的枝晶生长。不同的应力还会互相影响，例如金属迁移会因化学污染物和其构成成分而加速；因为热膨胀系数不匹配，热应力会激活机械失效。其他常见的例子还包括应力辅助腐蚀、应力腐蚀开裂、应力辅助扩散排空、场致金属迁移和温度引起的相关扩散加速现象，如金属间增长的动力、蠕变、腐蚀和相互扩散等。

表6.1中的过应力失效类型还包括由设计失误引发的硬件性能失效，但它不会引发不可逆转的材料损伤。此类失效包括由机械、热学和电学性能的不足引发的性能失效。过弹性形变和场载荷下结构的错误阻尼就是典型的机械设计失效。温度增加是一种热设计失效，它由关键热传导路径的过度热阻抗引发。电气设计失效包括不合理的屏蔽造成的电磁干扰、不合理的阻抗、电容设计引起的错误瞬时电压等。为方便起见，由表面性质的失效也包括在过应力失效类型中。

要设计出能承受载荷且不发生失效的硬件产品，设计师就必须了解所有的失效机理。失效机理和相关的模型对于设计试验、用来审查标称设计和制造规范的筛选以及由制造和材料参数中的过度异变引入的缺陷级别都很重要。本章给出了一些重要失效机理的简要说明，首先讨论了由性能不足引起的失效，然后是由不可逆的材料损伤造成的失效。

6.2.1 异常的机械性能

产品对于机械过应力载荷产生的异常反应可能会影响产品的性能，但它不一定会引发不可逆的材料损伤。此类失效包括机械静载荷引发的异常弹性形变、动载荷引发的异常瞬时表现（如固有频率或阻尼的改变）、与时间相关（粘弹性）的异常表现等。

例如分子键由于机械载荷拉伸而发生的弹性形变就是完全不可逆的，也就是说，载荷移除时，形变消失。一些时候，过应力载荷会引发功能性失效，如大型光学设备中，精密镜片的过度变形；大型、灵活的空间结构的巨大形变会触发不稳定的动态模式；电子设备中的连接线的交错、包装盖板或柔性电路的过度弯曲都会引发短暂的和（或）过度的干扰。

过度弹性形变的过应力损伤模型（Overstress Damage Model）通常涉及大变形理论（Large Deformation Theory），它建立在有限变形弹性（Finite-Deformation Elasticity）的非线性理论 [Malvern，1969] 之上。过应力损伤模型由非线性应变变形定义，它通过弹性本构关系，把应变与相应的应力关联起来。如果要求所关联应力与应变完全相等，就会产生一个非线性边界值问题，对于未知位移，由于受到合适的边界条件影响，此问题是可以解决的。由于存在一些几何或公差限制，当形变到达某个极限值时，就会引起过度弹性变形。其他文献资料对量化模型的细节讨论、案例有所描述 [Dasgupta and Hu，1992]。

6.2.2　异常的热学性能

热性能失效（Thermal Performance Failure）的出现是由组件中不当的热路径设计造成的。此情况包括单个部件的异常传导性和表面散热性，还包括传热路径的错误对流路径。不当的热设计引发的失效一般表现为部件运行过冷或过热，这样会导致运行参数逸出规范值。通常，基于冷却的性能恶化是可逆的。此类失效产生的原因包括直接热载荷和由机械载荷引起的热载荷，例如机械运动中的材料摩擦和耗散损失。另外，由电阻载荷引起的热载荷也是原因之一，它会反过来产生过度的局部热应力。充分的设计检验要求对热应力进行适当的分析，分析内容应该包括导电性、对流性和热辐射路径。

6.2.3　异常的电学性能

不当的电阻、阻抗、电压、电流、电容、电介体性能、不足的电磁干扰（Electro Magnetic Interference，EMI）屏蔽、离子辐射和静电释放（ESD）等都会引起电学性能失效。可将此类失效模式描述为电学参数的可逆漂流和（或）伴随的热故障。在此，我们仅讨论两种主要的电设计失效，即由不当的电磁干扰屏蔽引起的失效和由离子辐射引起的失效。

1. 电磁干扰

所有的电磁波都由磁场（H）和电场（E）组成，两者的比值取决于放射源的性质以及放射源与屏蔽设施的接近程度。E 对 H 的比率称为波阻抗，不考虑放射源类型，距离放射源非常远时其值为 1。此时可以称波为平面波，其阻抗为 377Ω（自由空间的波阻抗值）。相对于电压来说，如果放射源具有非常大的电流（如由变压器或输电线产生的电流），我们就称其为电流源、磁源或低阻抗源；如果放射源在高电压低电流下运行，那么称其为高阻抗源，此时把它的波作为一个电场看待。

很多电子电路都对电磁辐射非常敏感，需要对辐射进行屏蔽，以保证产品正常运行。电磁干扰和射频干扰（Radio Frequency Interference，RFI）都是电磁辐射的表现形式，EMI（电磁干扰）包括所有频率的波，但 RFI 只是那些频率高于 20kHz 的波。RFI

被定义为高频、高阻抗辐射电磁波，电场作为主要波组成，在其中占据主导地位。从直流电（Direct Current，DC）到20～25kHz范围是低频、低阻抗区域，此处，磁场支配EMI波。

RFI范围内产生的EMI放射主要是由高频率数字电路（如信号钟和高速逻辑器件）、无线电路和微波电路引起的。在电子产品中，如果屏蔽不当，此类辐射可能会影响临近电路或其他产品电路的运行，也可能会穿透箱柜，影响临近箱柜内电路元器件的运行。低阻抗EMI由变压器、电动机、螺线管、永久磁铁、电磁铁或其他产生外部磁场的电驱动设备产生。当EMI足够强时，此区域就会产生电流，进而影响其他部件的运行。

当电磁波遭到中断（如金属屏蔽物阻隔）时，如果波阻抗的大小与屏蔽物固有阻抗差异较大，大部分能量会被反射回来，只有很小部分能穿过界或被吸收。金属因其本身的高导电率而具有低阻抗性。对于低阻抗波来说，因为金属屏蔽物的阻抗非常接近于波的阻抗，所以波的能量中有少部分被反射回来，大部分被吸收。

2. 粒子辐射

电学失效模式由辐射引起，它在一定程度上规定了包装材料的选择以及包装材料中可允许的杂质，因此它对硬件设计非常重要。辐射的屏蔽对于包装设计和布局也非常重要。辐射对微电子产生了严重的影响，进而阻碍了大规模集成电路（Very Large Scale Integration，VSLI）密度的进一步快速增长。这些影响对存储芯片尤为重要，也导致了其他微电子技术的快速发展。粒子辐射也会引起材料的老化，我们将在磨损失效机理中对其进行深入讨论。现在，大多数微电子、宇宙射线或放射性污染物都会产生单粒子，它是因为单一高能粒子如电子、介子、μ介子、π介子或α粒子穿过了微电路而产生的。

单粒子翻转（软错误）是一种逻辑状态短暂、不可再现的改变，也就是说，它和物理损伤无关，失效字节会在下一次写入循环中恢复。从这个意义上讲，我们可以把单粒子翻转认为是电设计失效（见第6.2.3节）。粒子穿过微电路时会损失能量，从而导致直接电离。单粒子翻转是由粒子的直接电离产生的，它也能由中能核反应产生的次级粒子的电离产生，电离可产生电子——空穴对（Electron-Hole Pairs）。如果此电荷在加有反偏压的PN结附近产生，其强电场的存在会引起电子和空穴的分离，有适当标记的电荷被收集，有相反标记的电荷被迫移出耗尽层。

6.2.4 屈服

这是本章第一次讨论材料的过应力失效。在机械载荷超过屈服强度（有时称为流动应力）时，材料中会有微观结构缺陷（位错）发生。它所引起的塑性形变是不可逆的，也就是说，这是材料的永久性形变，即便载荷移除，形变也将持续存在。从功能上来说，这种永久性形变是不允许出现的，在一些硬件中，我们可以将其认为是过应力失效机理。常见的屈服包括：具有精细结构的光学器具座、计量设备和涡轮叶片等产品中的过应力塑性形变。

一些金属中不会出现超过屈服强度的机械硬化，即流动应力不会随应变力增加，进

而超过材料本身的弹性极限。一些有重要意义的机械硬化由 Ramberg-Osgood 幂定律［Hertzberg，1989］计算：

$$\sigma = K\varepsilon_{p}^{n_{m}} \tag{6.1}$$

其中，σ 为应力，ε_{p} 为塑性应变，n_{m} 是材料的应变硬化指数，K 是材料的塑性系数。这些材料常数的值可在美国材料协会（ASM）介绍大部分工程材料的手册中找到。对于塑性形变模型的定量讨论和案例研究，可以在文献［Dasgupta and Hu，1992］中找到。

6.2.5 扭曲

扭曲是一种过应力失效机理。在受到压缩载荷时，具有细长结构的零部件会突然出现剧烈的不稳定，这就引发了扭曲。扭曲失效的例子包括：细长柱受到轴向压缩产生的横向折叠、薄壁结构的横梁部分因弯曲引起的断裂、拥有薄壁的管状轴因扭曲产生的剪力屈曲、薄板和薄片平面内的压缩和剪切载荷引起的折皱等。当压缩载荷达到被称为屈曲临界应力（Critical Buckling Stress）的临界值时，就会引起不稳定。屈曲临界应力是材料特征（如硬度）和几何结构（如长度直径比）的函数。

从数学的观点来讲，扭曲是沿垂直于原变形方向，不稳定路径的形变可以用本征值（Eigenvalue）或分支理论（Bifurcation Theory）［Timoshenko and Gere，1961］来求解。凭借增量非线性算法（Incremental Nonlinear Algorithm），采用大变形理论可以完成对后屈曲（Postbuckling）的分析［Dasgupta and Haslach，1993］。

6.2.6 断裂

大部分材料都有局部小范围的瑕疵，如微尖裂纹（Sharp Microcrack）。在断裂出现之前，脆性材料在过应力载荷下会出现微小的屈服和不具弹性，过度应力集中在尖锐裂纹顶端会导致裂纹发生灾难性的扩散。在韧性材料中，重要的塑性区域可能会因为局部屈服而使裂纹尖端提前出现，使材料屈曲的能量会加速韧性材料对断裂的表观阻力。

脆性断裂设计（Designing For Brittle Fracture）是一门相对较新的学科，它起源于二战时期，因为当时盟军的船舰由钢板焊接而成，在冰冷的大西洋里变得很脆。现在，断裂被认为是造成工程硬件失效的主要原因，例如涡轮叶片、机身部件、桥梁、建筑框架、电子模型、玻璃和陶瓷部件等。准脆性（Quasi-Brittleness）会导致高硬度合金和陶瓷发生失效，热固性聚合物也会因为脆性断裂产生大范围的显微裂纹和龟裂。在其他韧性材料中（如焊锡），脆性金属间化合物的形成也会引发脆性断裂。基于应力的失效标准此时是不可用的，因为线性弹性分析不考虑远场平均值或公称应力，而只在瑕疵或裂纹的顶端分析无限应力，因此，需要用新的量度来量化应力场的严重性。这一参数被称为应力强度因子（Stress Intensity Factor），它指裂缝尖端应力场的强度。

Griffith 假设：当断裂固体上产生新的无裂纹表面所需的能量少于裂纹长度改变引起的应变能减少值时，灾难性裂纹扩散就会发生［Hertzberg，1989］。断裂机理所用的方法将用来推测断裂发生局部扩散时的远场应力级别。

应力强度因子 K 用来描述裂纹尖端应力场的强度，它是根据外施应力和裂缝尺寸来定义的。例如在长度为 $2h$、宽度为 $2b$ 的金属薄板中央有尺寸为 $2a$ 大小的裂缝，$a \ll b$

（表示薄板的尺寸为无限大），K_I 为

$$K_I = \sigma(\pi a)^{\frac{1}{2}} \tag{6.2}$$

其中，σ 是外施远场单轴应力。

裂纹即将扩散时，应力强度因子的阈值或关键值是材料抵抗脆性断裂阻力的分量，它被称为材料的断裂韧性（Fracture Toughness）。断裂韧性取决于裂纹相对于外施应力的方向，通常有三种不同的基本断裂模式：裂纹开放模式、剪切模式和撕裂模式。在ASM手册中可以找到常用工程材料的断裂韧性值。在断裂机理分析中，常用的设计方法是：假定零部件特征性缺陷发生在最高应力区域，计算关键远场载荷。详细的方法和例子可以在其他相关文献中找到［Dasgupta and Hu，1992］。

和脆性断裂一样，韧性断裂（Ductile Fracture）也是一种过应力失效机理。对它进行分析要求使用非线性建模方法，因为，当裂纹尖端存在大规模可塑性时，脆性断裂的线性弹性理论将变得不可用。韧性断裂在很多材料中出现，如铝、金、铜和锡，尤其是在高温情况下。材料在低温度、高应变率的情况下会出现脆性行为，在高温度和（或）高变形率的情况下会过渡到韧性行为。在韧性材料中，裂纹扩散需要高能量，因为裂纹尖端的非弹性形变会引起材料断裂韧性的明显增加。

在非线性情况下，应力强度因子不具实际的物理意义（除非是呈幂律分布硬化材料的裂纹，其特征是一个奇异场）［Broek，1978］，它也不再适用于表征断裂韧性。然而在Griffith的能量概念中，仍可以用它来度量裂纹扩散的能量需求，预测脆性断裂。本章只介绍基本概念，其他详细内容可以参考文献［Dasgupta and Hu，1992］。

最方便方法是以标量守恒积分来定义能量，它是用来定义裂纹延伸所需能量的能量动量张量（Energy-Momentum Tensor）的一部分［Broek，1978］。脆性材料中裂纹尖端扩散时，守恒积分也可用于表示的能量释放率，因为非弹性工作耗散的存在，它不适用于韧性材料。断裂韧性是守恒积分的关键阈值。常用的守恒积分是Rice的J积分［Broek，1978］。经过适当的修改，J积分适用于冲击载荷下的动态裂纹扩散、裂纹表面牵引（例如充满液体的裂纹）、裂纹尖端的大规模形变和蠕变应力松弛。

表述韧性材料对断裂抗性的另一种常见方法是裂纹尖端张开（Crack-Tip Opening Displacement，CTOD）。非弹性形变会使裂纹尖端的扩散变得迟缓，进而产生一个非奇异应力场，通常用等效尖裂纹的Dugdale-Barenblatt模型对其进行分析，也可以在屈服总量已知的情况下，利用塑性变形的滑移线理论（Slip-Line Theory）对其进行分析［Broek，1978］，也就是对CTOD的分析，我们认为断裂的关键值是材料抵抗断裂的特征值。

因为非弹性工作耗散的存在，对韧性材料中的稳定裂纹扩散进行设计非常艰难，这也是当前的主要研究领域之一。

6.2.7 接触面脱离粘连

在过应力载荷下，不同粘连材料间的接触面会发生粘连失效。这样的例子包括复合材料的脱层和粘合连接的粘连失效等。电子封装应用中常见的粘连失效包括芯片和其附着材料交接面处的粘连失效；接合线和接合盘的粘接失效；焊接连接中焊接材料和零件

材料之间的焊接点断裂等。

界面强度（Interfacial Strength）取决于界面的化学和机械性能。接触面粘连失效常出现在不同附着物间的扩散粘接、胶接、焊接和钎焊接头等连接形式中。导致两不同材料间界面粘合强度（Interfacial Adhesive Strength）提高的一个因素是材料的相互扩散，但是两种粘接材料各自不同的扩散特性却降低了界面强度（见第6.2.11 节）；同样，过度的金属间增长也会导致韧性不足的脆性界面。

界面粘合强度用来衡量在界面分离发生之前能穿过界面传递的最大机械功或能［Hertzberg，1989］。分离两材料间界面所需的功（或能）包括两个阶段：克服粘合能量所要做的功和使材料发生弹性或非弹性形变所需的功。因此，两脆性材料间界面的韧度低于两韧性材料间界面的韧度。总断裂能 G_f 为

$$G_f = W_a + W_p \tag{6.3}$$

其中，W_a 是可逆粘合功，W_p 是两阶段内不可逆非弹性形变的功。公式（6.3）忽略了两个阶段内使材料发生弹性形变所需的能，它小于 W_a 和 W_p。W_a 的定义为

$$W_a = \gamma_\alpha + \gamma_\beta + \gamma_{\alpha\beta} \tag{6.4}$$

其中，γ 是在接触面间建立新面所需的两表面间张力或表面张力，它被定义为所要建立表面的单位面积上 Gibb 自由能的改变率，它是摩尔单位材料在常温、常压条件下的度量值；下标 α 和 β 分别表示两个阶段的表面张力；$\alpha\beta$ 表示两阶段间的界面张力；W_p 是两阶段中不可逆形变的功。

一些经验主义者根据两材料间的电子结合能来描述界面结合强度，它是两材料间的唯一性质。在连续的范围内来看，机械强度是按照界面断裂韧性（Interfacial Fracture Toughness）来描述度量的。这是任何两种材料间界面的唯一特性，所以，必须要为硬件应用中常用的材料描述界面断裂韧性。

6.2.8 疲劳

这是本章讨论的第一种磨损失效机理。即便峰值应变力从来没有超过材料的韧性（失效处应变力）极限，材料中循环的机械变形（拉伸）和载荷（应力）最终也会引发失效。此类失效是由伴随载荷循环的增量损伤（Incremental Damage）的累积造成的，被称为疲劳［Sandor，1972］。疲劳属于磨损失效机理，包括裂纹的出现和扩散，它是工程硬件失效的主要原因之一。通常，裂纹延伸出现于材料的不连续处或者缺陷处，表现为局部应力或塑性应变集中。常见的例子是飞机机架和推进部件、土木工程结构（如桥梁和建筑）、金属化合物、电路板焊点中出现的疲劳断裂。

由于大应变振幅的存在，材料在 103 或者 104 载荷周期内会出现低循环疲劳（Low-Cycle Fatigue，LCF），由相对较低的应力振幅在大于 103 或 104 载荷周期内引发的失效，称作高循环疲劳（High-Cycle Fatigue，HCF）。材料的疲劳特征可由其应力—寿命（Stress-Life，S-N）曲线或应变—寿命（Strain-Life）曲线描述，这些曲线表示应力或应变振幅和不引发失效的载荷数量的相对关系。从以往经验来说，总应变振幅 $\Delta\varepsilon/2$ 与引发失效的循环次数 N_f 相关，Coffin-Manson 关系式描述了这一关系［Hertzberg，1989］：

$$\Delta\varepsilon/2 = [(\sigma_f/E_Y)(2N_f)^b + \varepsilon_f(2N_f)^c] \tag{6.5}$$

图6.1是这种关系的图形化表示。公式（6.5）右边的两部分分别由重对数图尺中的两条直线表示，第一部分表示高循环疲劳，第二部分表示低循环疲劳。其中，σ_f为疲劳强度系数（其相反数为引发断裂的实际应力）；E_y为杨氏模量；b为Basquin疲劳强度指数（图6.1中HCF线的斜率）；ε_f为疲劳延性系数（其相反数为引发失效的实际应变力）；c为Coffin－Manson延性指数（图6.1中LCF线的斜率）。b和c都与公式（6.1）中的循环应变硬化指数（Cyclic Strain Hardening Exponent）n相关。因此，材料的这几项参数完全可以描述材料的疲劳行为。

公式（6.5）可用于计算完全反向的循环载荷，如果载荷不是完全反向的，那么需要引入Foreman校正因子，用（$\sigma_f-\sigma_{mean}$）替换σ_f，σ_{mean}为外施循环应力的平均值。这种疲劳法也可用于变化的历史载荷，只要把载荷历史被划分为区段，那么每一区段的应力平均值和应力振幅都为常数。每段载荷引发的疲劳损伤可以分别计算出来，累积损伤就可由叠加模型如Miner疲劳损伤累积法［Sandor，1972］计算。

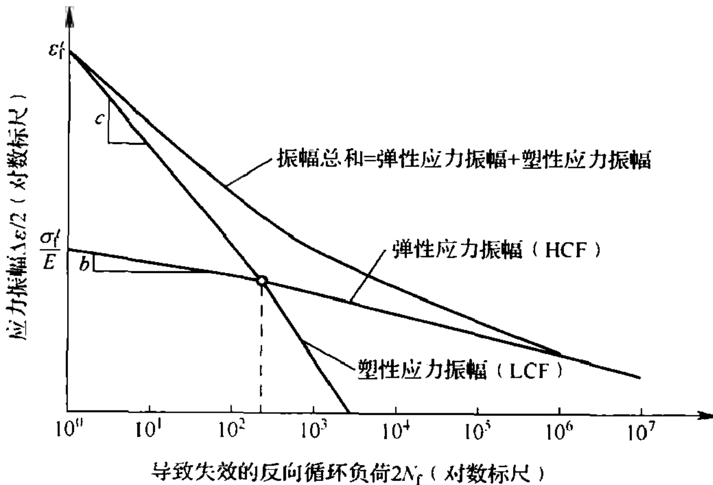


图6.1 Coffin-Manson关系

一旦出现疲劳裂纹，它就会在循环应力下稳定扩散，直到在外施应力振幅下变得稳定。疲劳裂纹的扩散是非常重要的阶段，因为它是最严重的疲劳裂纹处发生过应力失效的先兆。当裂纹扩散速度da/dN超过应力强度因子ΔK的临界循环范围时，就会变为零。最初，裂纹扩散由扩散速度定义，扩散速度随ΔK的增加而减小；在裂纹呈稳定状态的第二阶段，扩散速度与重对数图尺中的ΔK成线性比例关系；在第三和最后阶段，扩散速度急剧增加直到灾难性失效发生，如图6.1所示。对于设计目标来说，稳定状态（第二）阶段是最重要的。

对于周期性载荷，稳定状态的疲劳裂纹扩散模型由Paris幂法则定义：

$$\frac{da}{dN}=\frac{A_c\Delta K^{p_c}}{[(1-R_0)(K_{IC})-(K_{max})]}(R_0>0) \tag{6.6}$$

其中，A_c和p_c为材料常数。为了描述这些性质和很多工程材料的属性值，ASTM进行了一些标准试验，这些数据可以在ASM手册中找到。

R_o 由以下公式计算：

$$R_o = \frac{\sigma_{min}}{\sigma_{max}} \tag{6.7}$$

对公式（6.6）中的增长裂纹大小进行积分，可以得到裂纹由发生到扩散，再到引发失效的循环次数 N_p。缺陷在外施应力下会变得不稳定，在此之前，积分的范围为从初始尺寸 a_0（最小可检测瑕疵）到最大允许尺寸 a_1、a_2，即

$$a_f = \left(\frac{1}{\pi}\right)\left(\frac{K_c}{\sigma_{max}}\right)^2 \tag{6.8}$$

其中，K_c 是材料的断裂韧性。

6.2.9　蠕变

在高温环境中的机械应力下，一些材料如热塑性聚合物、焊锡还有其他金属材料会发生与时间相关的形变，称为蠕变。实际上，很多的形变都发生在有限时间段之内。为了方便力学建模，我们把很短时间段内发生的形变当做“瞬时形变”，根据形变的可恢复程度，称其为弹性形变或塑性形变；把很长时间段内才会发生的形变称为“蠕变”，根据形变是否能恢复，可分为粘弹性（滞弹性）形变或粘塑性形变。

蠕变是一种引发功能性失效的磨损失效机理，功能性失效出现的原因是过度形变或者呈现为蠕变断裂前兆的形变。蠕变出现的原因是位错攀移机制、高分子链调整、晶界滑动（超塑性）、晶内无效迁移（自扩散）和（或）晶内或穿晶无效迁移（晶界扩散）。在不同温度下，同一材料中会出现不同的占主导地位的蠕变机理，有时，多种蠕变机理会同时出现。

很多材料都存在蠕变率下降的阶段（初始蠕变），然后是平稳阶段（第二阶段蠕变），在最后阶段（第三阶段），蠕变率会增加。在电子封装的寿命内，设计师必须确保蠕变应变在设计约束内。在适当的应力级别下，由平稳阶段（第二阶段）蠕变产生的蠕变应变值 ε_c，通常由 Weertman 蠕变法则［Hertzberg，1989］表示为

$$\varepsilon_c = C_1 s^{n_1} t \exp\left(\frac{-E_a}{K_B T}\right) \tag{6.9}$$

式中　t——已消逝时间；

s——应力；

K_B——Boltzman 常数；

T——热力温度；

E_a——蠕变机理的激活能。

相关材料的 C_1、n_1 和 E_a 都从试验数据而来，很多常见材料的这些值都可以从工程手册中找到。其他细节和案例可参阅文献［Li and Dasgupta，1993］。

6.2.10　磨损

磨损是一种磨损失效机理，对于所有受外部微粒或接触表面滑动影响的硬件而言，它是极其重要的。例如沙粒、水或其他材料外部微粒的冲撞会导致逐渐侵蚀，进而引发磨料磨损；轮齿、滑动轴承表面、活塞和气缸组件等之间会出现摩擦磨损，进而引发粘

附磨蚀；在液体管道中，冷却管道内的气蚀会引起液体侵蚀，进而会引发磨损。粘附磨损可导致点蚀和粘结现象。磨损不仅是一种失效机理，它还导致硬件对随后的腐蚀和过应力失效的抗性降低。

滑动接触面间的粘连磨损一般被描述为

$$V = k_m Fvt \tag{6.10}$$

式中 V——材料的去除量；

F——触点压力；

v——两装配面的相对滑动速率；

t——持续时间；

k_m——一种材料特征。

详细阐述和案例参见文献［Engel，1993］。

6.2.11 相互扩散引起的老化

当两种材料紧密结合时，一种材料的分子会通过扩散转移到另外一种材料中去，材料原子的运动也能引发此类现象。从原子的观点来看，扩散是晶格节点处的原子迁移现象。原子必须拥有足以破坏现有节点的能量，才能在另外的晶格处形成新的节点。扩散速度是一种典型的材料特征，它可以通过实验度量获得。扩散过程的最常见数学描述是Fick第二定律：

$$\frac{\partial C_d}{\partial t} = \nabla^2 (D_d C_d) \tag{6.11}$$

其中，C_d 是扩散物种的浓度，∇是梯度算子空间，t 是时间，D_d 是扩散系数。可以通过明确定义的实验室试验描述 D_d 的值，大部分工程材料的扩散系数可以从材料手册中查得。

扩散现象本身不是一种失效机理，例如对于扩散粘合来说，它还是有益的机制。然而，当扩散介质是有害物质或是化学腐蚀品，那么它就表现为一种失效动因；另外，当扩散导致微观结构的老化、有害的蠕变变形、金属的迁移和不平衡的相互扩散等现象时，它也是一种失效动因。

当两种材料必须相互粘合时，相互扩散现象对于形成接触面粘合力是非常重要的。然而，如果两种材料的有效扩散速度不相等，其中一种材料就会耗尽其原子，从而引发可肯达尔效应（Kirkendall Effect），并损失接触面处的强度。常见的例子是电子设备丝焊点处金浸出到铝中引起紫斑（Au-Al 合金加热缺陷）。两种粘合材料的互扩散常数一定要相近，这样才能避免此类失效机理。换句话说，某段时间内，由温度协助的过度扩散会使界面处出现金属化合物的增加，进而导致脆性界面韧性不足。

相互扩散是一种与时间相关的现象，因此，它是一种磨损机理。其他详细资料和实例参见文献［Li and Dasgupta，1994］。

6.2.12 离子辐射引起的老化

离子辐射是宇宙空间、建立在地面的核能和离子研究基地中的一种常见现象。辐射的危害同时包括机械失效和电学失效。机械失效是典型的脆性老化现象，属于磨损失效

类型。常见的例子包括太空卫星和反应堆容器中的外露硬件发生的失效。电学失效是一种过应力现象，即由单个辐射离子穿过（超）大规模集成电路（LSI/VLSI）引起软失效的现象（见第6.2.3节）。

不同材料中，辐射损伤会引起不同类型的老化。辐射损伤是一种与时间相关的磨损现象，它常出现于金属、陶瓷和聚合物材料中。在金属和陶瓷中，辐射会引起点缺陷（Point Defect），例如晶格结构分子中撞击出的原子暂留在间隙处，从而形成的成对原子空缺和间隙（Schottky 缺陷）。这类点缺陷会导致脆性老化，退火工序可以解决此问题。更严重的是在电子封装应用中，这类缺陷还能引起热、光和电特性的改变，从而损害正在运行的设备。在聚合类材料中，聚合链的断裂、由聚合链分支引起的聚合程度改变都会引起辐射老化。这两种现象都能使聚合物的强度降低，最常见的现象是聚合物长时间暴露在强紫外线下会发生光降解。有时，稳定剂会对这类磨损失效有所抑制。

6.2.13 其他老化现象

随着时间的推移，其他各种形式的老化都会改变材料的性能。这样的例子包括：材料的氢脆性，热致解聚，热固化聚合物中增加的交联导致的脆性，晶体材料中的晶粒增长。这些机理的详细论述都不在本章的讨论范围内。

6.2.14 腐蚀

腐蚀由金属的化学或电化学降解引起。它是一种与时间相关的磨损失效机理，通过脆性断裂表现为过应力失效的前兆；通过疲劳裂纹扩散表现为磨损失效的前兆。在微观范围内，腐蚀还能改变材料的电和热行为，常见的三种腐蚀是均匀腐蚀、电解腐蚀和点状腐蚀。腐蚀反应速度取决于材料本身、存在的电解质、存在的离子污染物、几何因素和局部电偏压（Local Electrical Bias）等因素。

均匀腐蚀是一种均匀的化学反应，它发生在金属电解质界面的表面。腐蚀过程的连续性和速度取决于腐蚀产物的自然特性。如果腐蚀产物是可溶于电解质的（例如水），那么它会被溶解掉，新的金属暴露出来，进一步被腐蚀；如果腐蚀产物是不可溶的、无细孔的粘附层，那么它将影响反应的速度，最终导致腐蚀停止。

当两种或更多种金属相互接触时，会出现电解腐蚀。每种金属都有其独特的电化学势，当两种金属相接触时，具有高电势的金属成为阴极，另外一种金属则为阳极；当两种特性不同的金属间发生电接触，就会形成原电池。电解腐蚀的速度由阳极的电离速度决定（也就是阳极材料处，离子传递到电解液的速度），电离速度由两接触材料间的电化学势决定。腐蚀介质的电传导性影响着电解的速度和分布。在具有高传导性的溶液中，较活跃金属的腐蚀会散布在较大范围内；而在大部分低传导性的溶液中，更多的电解发生在两不同材料间的电接触点附近。

点状腐蚀发生在固定的区域，它会导致材料表面形成凹坑。凹坑内部的腐蚀条件会使腐蚀的速度加快。正极处的阳离子在流向溶液的过程中会被水解掉，然后产生氢离子。这会导致凹坑处的酸性增强，进而破坏掉附着在凹坑的腐蚀产物，将新的材料暴露在电解环境内。由于凹坑内的氧元素供应较少，还原反应仅会在凹坑的出口处发生，这也限制了凹坑的横向生长［Pecht and Ko，1990］。

表面氧化是另外一种常见于金属材料的腐蚀，它由形成氧化物的自由能决定。例如铝和镁的氧化物有很强的驱动性，更不用说铜、铬和镍的氧化物了。根据腐蚀反应的化学计量法，氧化物形成的类型可被分为可渗透性氧化物和稠密型氧化物。氧化类型通常决定了后续的腐蚀速度。厚的、无细孔的氧化层会形成保护层，切断氧气的供应防止表面上进一步氧化的发生，例如铝和不锈钢。有时，剥落的氧化层中的腐蚀产物（氧气化物）的量会多于材料本身的量。这种结垢失效（Scaling Failure）会将底层金属暴露在新的腐蚀环境中。

腐蚀是工程硬件损伤和失效的主要原因之一。防止腐蚀、处理腐蚀产物、新老材料的更替都要花费很大的财力。

6.2.15 金属迁移

此类磨损失效机理由扩散现象驱动，对电子硬件非常重要。金属迁移有很多种类型，包括电子迁移、阴极的枝晶生长（Dendritic Growth）和阳极的导电纤维（Conductive Anodic Filament，CAF）增长。枝晶生长的本质是一种电解过程。在此过程中，材料由阳极区迁移至阴极区。金属迁移会在连接区导致连接漏电，如果是完全连接，则会引起短路（迁移性电阻短路）。虽然关于银的迁移曾受到广泛关注，但其他电子金属，如铅、锡、镍、金和铜也都会发生迁移，这取决于环境条件。由于和时间相关，此类现象也是一种磨损失效机理。

金属迁移由金属本身的电子性、电解质（如冷凝水和离子种类）和存在的电压差异决定。需要对金属迁移敏感的材料进行水蒸气和离子污染物隔离。因为迁移现象是一种电解过程，所以导电介质是必不可少的。可导电的离子形式包含有杂质，如氯化物和腐蚀产物。引发金属迁移所需的驱动力是电子类产品在偏压条件下的潜在差异。虽然此类失效的主要应力是电位梯度，但它会因为次级应力如湿度、离子污染物和温度而加速。

6.3 载荷、应力和材料行为

硬件在制造、装配、试验、返工、维修、运输、仓储、搬运和运行的过程中，会承受很多载荷。损伤可能在任何一个环节出现并累积，从而影响运行过程中硬件产品的可靠性。因此，在某些环节中，量化相关应力是非常有必要的，这样有利于可靠性的理解和建模。

应力是引发第6.2节中所讨论失效机理的促进因素。为了便于阐述，我们把应力定义为衡量失效位置处载荷分布密度的局部分量，例如循环机械力加载在一个孔的局部会导致局部应力集中，从而引起高循环疲劳裂纹开始出现。在此情况下，外施远场循环机械力为载荷。为了得到载荷引发的应力，我们需要分析零部件的几何形状、材料的结构性质和部件的周边条件。一般来讲，这涉及到初始边界值的求解。对部件几何特性和线性材料特征进行封闭分析，就可以进行应力分析了。然而对于其他复杂情况或速度相关的非线性材料特征，需要用近似数值计算方法来分析，如有限差分法、有限元法、边界

元法或其他方法。精确的应力分析是可靠性建模中的一个关键性步骤。

如第 6.2 节中所阐述的那样，引起应力的载荷类型包括：电流和电压的等级及其梯度、污染物浓度及其梯度、温度梯度、湿度、机械静力、动态振动和冲击载荷、电磁场和粒子辐射。运行应力通常作用在硬件元素之上，它会驱动各种失效机理，进而导致零部件失效。特定零部件的可靠性由其运行环境和主要的、导致零部件失效的失效机理所决定，运行应力对它来说非常重要。在潜在失效位置处，需要为所有硬件元素确定关键应力，并要根据关键应力的相对严重性对其分级。

实际上，某物理量是应力还是载荷，取决于它在失效过程中所扮演的角色。例如在设备过度高温（燃烧）的失效机理中，温度是一种应力，它同时出现于损伤模型和加速形变方程中；然而，在由温度引起的半导体芯片开裂过程中，温度却是一种载荷，因为它会引起机械应力［Hu，1994］。按照惯例，在电子封装中，为获取温度级别和分布而进行的应力分析被称为热分析，为获取由温度引起的机械应力的级别和分布而进行的应力分析被称为热应力分析。这些分析的目的是在已知装配体材料和几何形状条件下，研究外施载荷和应力的相互关系。每一位置的应力都是其载荷的函数，它包括产品部件和模块的各种温度参数、几何特征和材料特性。

良好的可靠性预测还需要科学的量化技术，以精确地定义产品需求和设计运行环境需求。生命周期应用剖面是一个有序的时间列表，它列出了所有可能引起失效的载荷。这些载荷组成了量化已知应用条件的参数。例如硬件设计师必须在特定的环节记录飞行应用，包括发动机预热、飞机滑行、攀爬、巡回、高速飞行演习、开火、弹道影响、速降、急停等信息。必须要将它们与适当的运行载荷如加速度、振动、冲击力、温度、湿度和电力循环等联系起来，否则，这些信息对硬件设计师毫无用处。

在所有决定可靠性的关键性参数中，相关材料组成和在制造、加速试验、仓储、搬运和运行过程中超过所有预期载荷范围的损伤特征是最重要的。为了获得基于物理的可靠性预测，我们必须用合适的随机分布函数来描述这些材料特征及其相应变化。不幸的是，关于材料性质的信息总是严重不足，这将使可靠性预测的精度大打折扣。各种材料的特征可以在 ASM 国际手册中找到。通常，制造商会因为相关材料修订自己的材料数据库。充足的材料知识是非常重要的，这一点怎么强调都不过分。每一个可靠性预测小组都应该查找足够的资源，以收集、编撰并传播关于基础材料性质的资料。

6.4 变异性与可靠性

在 6.3 节中，我们列出了所有失效/损伤模型的参数，这些参数包括材料性质、外施应力和失效标准。反过来，应力取决于载荷、几何特征、边界条件和材料组成特征。如果所有这些参数都可建模为决定性变量，那么我们就可以量化为所有过应力失效机理预测的失效强度和为磨损失效机理预测的失效时间。

然而，现实中还需要对所有参数的变异性进行建模。概率分布的估计均值和标准偏差可以完成此任务。例如材料构成和损伤特征通常会因微观结构和工艺的变化而发生变

异；几何参数通常会因为特定公差范围而产生变异；应用条件会因载荷历史而产生变异。因此我们认为，在特定应力和时间处，最好的硬件设计师通常能够基于适当的失效机理预测失效出现的概率，此概率表示设备在已知时间和应用情况下的不可靠性。下一节我们将介绍一种计算可靠性的方法，并结合案例进行说明。

6.5　可靠性预测技术

就像在本书的前言中所提到的一样，基于失效物理的可靠性预测对于设计和相关的决策，如生命周期后勤保障和成本评估来说是非常重要的。由于缺乏适当的失效机理模型和数据，可靠性预测技术通常都具有一定的主观性。因此在每次预测之后，用鉴定试验对结果进行验证是非常有必要的。本节将介绍一些非常适合与失效物理概念相结合的可靠性预测方案，然后讨论加速鉴定试验方法以及与后勤方面的可靠性预测。

对可靠性评估的过应力失效进行概率建模的一种可行的办法是应力—强度干涉理论（Stress-Strength Interference Theory）（详细内容和案例参见文献［Kapur and Lamberson，1977］等）。

此方法利用具体指定的概率密度函数（Probability Density Function，PDF）把外施应力描述为连续随机变量。基于包括几何公差和材料组成性质对失效位置影响的应力分析，尽可能好地估算应用剖面中所定义载荷的平均偏差和标准偏差。同样，基于失效位置处材料的平均和分散损伤特征，把强度假定为一个用概率密度函数表示的连续随机变量。现实中，获得这些概率密度函数的精确平均偏差和标准偏差是非常困难的。

图6.2分别描绘出了应力和强度的概率密度函数：$f_s(S)$ 为应力的概率密度函数，$f_\sigma(\sigma)$ 为强度的概率密度函数。可靠度 R 为强度 S 超过应力 s 的概率，分布中所有可能的应力值由公式（6.12）计算［Kapur and Lamberson，1977］：

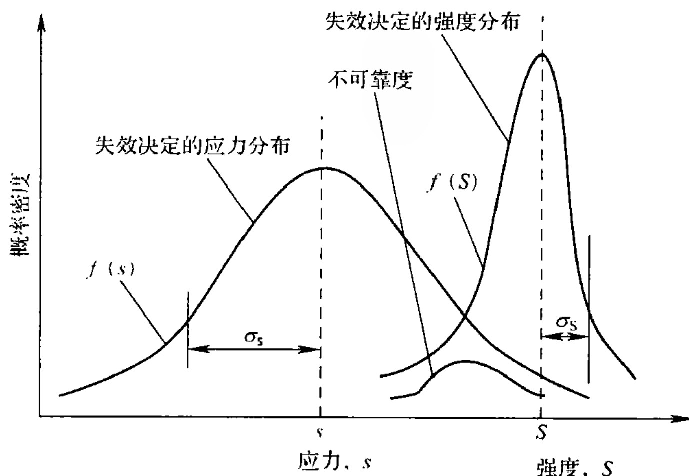


图6.2　结构化可靠性规划的基本构造

$$R = \int_{-\infty}^{\infty} f_s(S)\left[\int_{s}^{\infty} f_s(S)\,\mathrm{d}S\right]\mathrm{d}s \tag{6.12}$$

或者 R 为强度 S 的值超过应力 s 的概率，对于所有的 S，R 由式（6.13）计算：

$$R = \int_{-\infty}^{\infty} f_s(S)\left[\int_{-\infty}^{S} f_s(s)\,\mathrm{d}s\right]\mathrm{d}S \tag{6.13}$$

不可靠度 Q 可以由 $R=1-Q$ 获得。

或者定义一个新函数 $Y=S-s$，可靠性 R 是 Y 值为正的概率。如果 S 和 s 统计独立，那么：

$$R = \int_{0}^{\infty} f_y(y)\,\mathrm{d}y \tag{6.14}$$

实际中，获得精确概率函数比较困难，这正是应力—强度积分方法的局限性所在。可靠性预测的精度取决于最好置信度处的概率密度函数尾部的精度。为概率分布函数使用极值分布，可以在一定程度上缓解此类问题。

通常，应力—强度干涉理论可用于过应力失效机理问题，而对于磨损失效机理问题，所使用的概率方法稍有不同。其中一种方法是使用时间界面模型（Time Interface Mode，有时被称为强度退化模型，见第 4 章），另外一种方法是使用一个单独的无量纲损伤曲线［在图 6.3 中记为 $f_d(D)$］来代替应力和强度分布。损伤参数 DM 与时间相关，它随消耗时间（在疲劳载荷中，为已用载荷周期）而单调增长。根据应力历史，可以对 DM 的瞬时值与损伤参数和损伤累积进行适当的定义。例如在高循环疲劳损伤累积的情况下，按照惯例，在外施应力幅为 $\Delta\sigma$ 时，根据 Coffin-Manson 法则［公式(6.5)］，把每载荷周期的损伤参数 DM 定义为平均疲劳寿命 N_f 的倒数。在 Coffin－Manson 法则中的 HCF 情况下，随机建模技术用来表示决定外施应力幅度 $\Delta\sigma$、材料损伤常数中的不确定因素 Δf 和 b 的损伤曲线参数中的不确定因素。

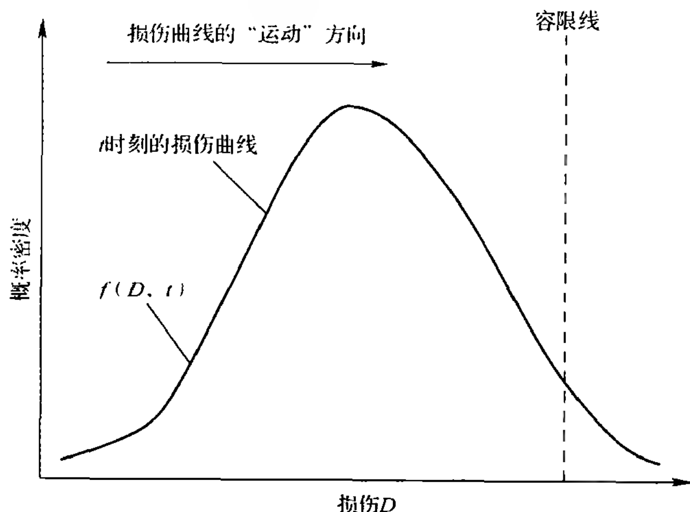


图 6.3　应力和强度的分布

在实际中，确定损伤曲线的平均偏差和标准偏差的精确值是极其困难的。Miner 法则［Sandor，1972］提供了一个简单的线性损伤叠加法，它用于计算复杂载荷历史下的损伤累积。在不存在愈合现象（如韧化）的简单材料中，损伤曲线随着损伤累积的增加向右单调迁移。因此，疲劳载荷历史的任何时刻都有唯一的损伤曲线。简单来讲，分析员一般认为平均值 μ_D 是时间的一个函数，而认为 DM 的方差是时间恒定的。在这种无量纲损伤范围内，可靠性 R 由 $DM>1$ 的概率得到。因此，$t=t^*$ 时刻的可靠度为

$$R(t=t^*)=\int_{1}[f_D(x)]_{t=t}*\mathrm{d}x \tag{6.15}$$

我们曾在第 4 章讨论过其他可靠性预测模型，其他各种参考文献［Haugen，1980；Kapur and Lamberson，1977；Lewis，1987］中也曾讨论过这些模型。一般来说，可靠性计算涉及到多随机变量函数的处理。这些都可以用近似形式完成，但我们可能需要最简单的数值结构方法，如 Monte-Carlo 方法。另一种比较新的技术——随机有限元分析也能提供数值化解决方法［Ghanem and Spanos，1991］。

下一节将描述一个案例研究，它使用了简单的闭合应力分析。在此案例中，我们确定了电子设备丝焊组装中的失效机理，并给出了用于预测可靠性的相关模型。

6.6 案例研究：微电子封装中的丝焊组装

这个简单的案例研究描述了一种失效物理概率方法在可靠性预测和模型中的应用过程。案例硬件是一个微电子封装中的焊丝以及丝焊组装。在此案例中，我们采用量化模型确定了潜在失效机理。在考虑了材料特性中的样本变异的情况下，使用材料强度方法分析了由于热循环载荷产生的热机械应力，并为一个预期疲劳寿命为 10000 的热循环估计了可靠度。此处仅给出了分析的重要过程，详细内容见参考文献［Hu，Pecht and Dasgupta，1991］。

6.6.1 失效机理和应力分析

在热循环情况下，丝焊存在线材的重复屈曲、焊盘和金属丝间的重复剪切应力、焊盘和基片间的重复剪切应力、金属丝的重复轴向应力［Pecht，Dasgupta，and Lall，1989］，这些因素造成的主要后果是疲劳失效。为了更好地为这些失效机理建模，我们要获取应变循环和热循环的关系，以便得到硬件产品的疲劳寿命和可靠度。

1. 金属丝屈曲

金属丝会随着温度的改变而膨胀或收缩，因而丝焊会经历弯曲疲劳。金属丝和基片的不同热膨胀是由温度循环引起的金属丝弯曲而产生的，它会导致楔形焊接和针脚焊接中的焊点根部产生应力反向，最终会导致金属丝的疲劳失效。图 6.4 描绘了热循环引起的楔形焊点变形。因为金属丝的横断面在楔形和针脚焊接的焊点附近减小，应力在根部集中，此处就可能成为一个由于金属丝弯曲而引起失效的点。

由于热膨胀的存在，金属丝会发生纯弯曲，我们把楔形焊接金属丝建模为一个纯弯曲下的横梁。温度循环中的热膨胀会在金属丝脚部形成弯曲应力，此弯曲应力的循环振

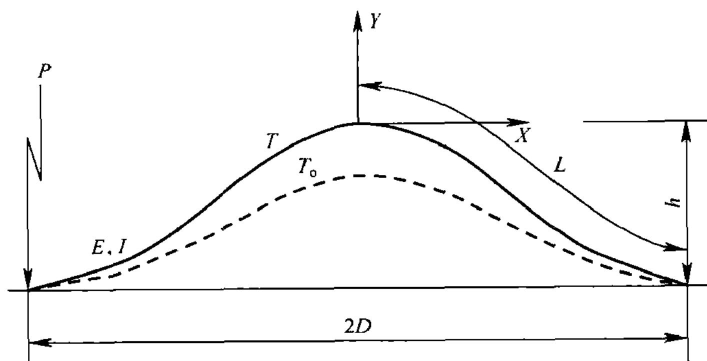


图 6.4 由热循环引起的楔形焊点变形

幅 ΔT 可用梁的弯曲理论和简单的线性弹性理论计算为（参见文献 [Hu et al，1991]）

$$\sigma = 6E_{w}\frac{r}{D_{s}}\left(\frac{L}{D_{s}}-1\right)^{\frac{1}{2}}\left(2\alpha_{s}+\frac{\alpha_{s}-\alpha_{w}}{(1-D_{s}/L)}\right)\Delta T \tag{6.16}$$

式中 α_w 和 α_s——金属丝和基片材料的热膨胀系数；

E_w——金属丝材料的杨氏模量；

L——金属丝的长度；

D_s——金属丝两端焊点的间距。

导致失效的屈曲循环次数 N_f 与用 Coffin-Manson 公式（见第 6.2.8 节）计算的疲劳应力范围相关：

$$N_f = C_w\sigma^{-m_w} \tag{6.17}$$

其中，σ 由公式（6.16）计算，C_w 和 m_w 是疲劳特征，由金属丝材料的拉伸疲劳试验决定。

2. 焊盘的剪切应变

随着温度的改变，双金属焊接会承受剪切应力，这是由金属丝和焊盘或焊盘（Bond Pad）和基片（Substrate）之间的不同热膨胀引起的，如图 6.4 所示。因为与基片相比，焊盘是典型的特别薄件，我们把它建模为金属丝和基片间的夹层。忽略掉所有弯曲变形，焊盘的剪切应力分布的一阶近似值可用剪滞模型 [Shear Lag Model，Jones，1975] 计算为

$$\tau = \frac{G_p\Delta T}{b_pZ}\left\{(\alpha_w-\alpha_s)-\frac{(\alpha_s-\alpha_p)}{1+(E_sA_s)/[E_pA_p(1-v_s)]}\right\}\frac{\sinh(Zx)}{\cosh(Zl_w)} \tag{6.18}$$

式中 G——剪切模量；

E——杨氏模量；

v——Poisson 比；

α——热膨胀系数；

脚注 p——焊盘；

s——基片；

w——指金属丝；

b_p——焊盘厚度；

A_p——焊盘的横断面积；

l_w——金属丝长度；

Z——依据材料特征值和丝焊组装几何形状表述的本征值［Hu et al，1991］。

公式（6.18）表明，焊盘的剪切应力是沿中间层位置的最大值 $x=l_w$ 的函数，它沿粘合剂—焊盘金属丝中心方向或粘合剂—焊盘基片界面中心方向减小。注意：$Zl_w \gg 1$。因此，在关键点 $x=l_w$ 处，$h(Zl_w)$ 的正切值，也就是温度循环导致的最大剪切应力振幅 ΔT 为

$$\tau_{max}=Q\Delta T \tag{6.19}$$

其中，

$$Q=\frac{G_p}{b_pZ}\left\{(\alpha_w-\alpha_s)-\frac{(\alpha_s-\alpha_p)}{1+(E_sA_s)/[E_pA_p(1-v_s)]}\right\} \tag{6.20}$$

一旦确定了剪切应力的最大振幅，可以通过 HCF 的 Coffin-Manson 公式预测导致粘合剂—焊盘材料出现剪切疲劳失效的循环次数为

$$N=C_{p'}\tau_{max}{}^{-m_{p'}} \tag{6.21}$$

其中，$C_{p'}$ 和 $m_{p'}$ 是通过实验决定的，它是粘合剂—焊盘材料的剪切疲劳特征值。常用工程材料的相关值可在工程手册中查得。

3. 金属丝和基片的剪切应变

使用与公式（6.18）中相似的参数，金属丝和基片的最大剪切应力可以计算为

$$\tau_{W_{max}}=\left\{\frac{r_w^2}{4Z^2A_w^2}\left[\frac{\cosh(Zx_w)}{\cosh(Zl_w)}-1\right]^2+\frac{Q^2\sinh^2(Zx_w)}{\cosh^2(Zl_w)}\right\}^{\frac{1}{2}}\Delta T \tag{6.22}$$

$$\tau_{S_{max}}=\left\{\left[\frac{W_pQ}{2ZA_s}\left(1-\frac{\cosh(Zx_s)}{\cosh(Zl_s)}\right)+\frac{(\alpha_s-\alpha_p)}{(1+v_s)/(E_s)+A_s/(E_pA_p)}\right]+Q^2\frac{\sinh^2(Zx_s)}{\cosh^2(Zl_s)}\right\}^{\frac{1}{2}}\Delta T \tag{6.23}$$

其中，

$$x_w=\pm\arctan h(A_w/r_w);$$

$$x_s=\pm\arctan h(A_s/r_s);$$

r_w——焊接金属丝的半径；

A_w——金属丝的横断面积。

公式（6.4）已定义了其他项。引起金属丝和基片材料剪切疲劳失效的循环次数也可以通过 Coffin-Manson 公式建模，即

$$N=C_{w'}(\tau_{W_{max}})^{-m_{w'}} \tag{6.24}$$

$$N=C_{s'}(\tau_{S_{max}})^{-m_{s'}} \tag{6.25}$$

其中，$C_{w'}$、$m_{w'}$、$C_{s'}$ 和 $m_{s'}$ 分别是金属丝材料和基片材料的剪切疲劳特征值，它们可以从实验室中的控制疲劳试验得到或查询工程手册获取。有了大部分材料特征，就一

定可以用测量平均值和标准偏差描述变异性。文献中的材料属性数据不足以获取标准偏差的真实估计值。

4. 线材的轴向拉力

在塑料封装中，密封剂包裹金属丝并与之接触，金属丝和密封剂之间不同的温度循环会导致它们产生不同程度的膨胀。温度的反复波动会引起金属丝的轴向疲劳。金属丝和密封剂轴向变形总量是由温度上升 ΔT 引起的变形加上它们所承受机械力引起的变形，相容性条件要求金属丝和密封剂的变形总量相等。使用此条件，再加上组装平衡，轴向应力的近似值为 [Hu et al, 1991]

$$\sigma_w = E_w(\alpha_e - \alpha_w)\Delta T \tag{6.26}$$

把公式（6.26）带入 Coffin-Manson 公式，引起失效的循环次数为

$$N = C_w \sigma_w^{-m_w} \tag{6.27}$$

其中，C_w 和 m_w 已在公式（6.17）中定义。

前面引用的五种疲劳失效机理同时出现。基于这些机理，每种失效模式的疲劳寿命 N_i 可以根据种类表述为

$$N_i = C_i S_i^{-m_i} \quad i = 1, \cdots, 5 \tag{6.28}$$

其中，N_i 是第 i 种失效机理中引起失效的循环次数。

$C_1 = C_5 = C_w$ 和 $m_1 = m_5 = m_w$ 都要在拉伸疲劳试验中确定；$C_2 = C_{p'}$、$C_3 = C_{w'}$、$C_4 = C_{s'}$和 $m_2 = m_{p'}$、$m_3 = m_{w'}$、$m_4 = m_{s'}$都要在剪切疲劳试验中确定。通过公式（6.15）可知 $S_1 = \sigma$，通过公式（6.18）可知 $S_2 = \zeta^{max}$，通过公式（6.21）可知 $S_3 = \zeta_w$，通过公式（6.22）可知 $S_4 = \zeta_s$，通过公式（6.26）可知 $S_5 = \sigma_w$。

这五种机理的任何组合形式都会引起丝焊的失效。假定这些机理是以串联形式出现的，如果 S_i、C_i 和 m_i（$i = 1, \cdots, 5$）的平均值已知，引起失效的循环次数平均值（对于每一种机理都是一个常数）可以通过公式（6.28）计算。与最短寿命对应的损伤机理或者是失效机理占有主导地位。在丝焊失效中，占主导地位的机理取决于运行条件、粘接材料的疲劳特征和制造过程中的粘接条件，它会因所需的可靠性级别而有所变化。

6.6.2　变异性和可靠性的随机建模

最常见的疲劳试验是在恒副应力下，用平均应力和应力范围定义的试验。即便是在非常仔细地控制试验条件的情况下，疲劳寿命数据也存在变异性。因此，足够的试验数据的完整描述应该包括生存函数（可靠性）R、应力范围 S_i 以及疲劳寿命 N 之间的相互关系。一般来讲，材料疲劳强度可用同时取决于应力级别和疲劳寿命的可靠性函数描述为

$$R = F(S_i, N) \tag{6.29}$$

对于已知应力级别，对数正态分布和 Weibull 分布通常用来做疲劳分析。度量平均寿命的参数 μ_N 分散在平均寿命周围；寿命的标准偏差 σ_N 可以通过在已知应力范围内，用极大似然估计法对试验数据估计得到。变异系数 $\delta_N = \sigma_N/\mu_N$，它是变异性的一个无量纲量。

使用对数正态分布，主要是出于数学计算的方便性考虑，但是它表示故障率函数降

低的概率，此函数与观测到的现象相矛盾。假定大范围独立事件会导致裂缝产生，对数正态分布与中心极限定理就可提供一定的理论依据。Weibull 分布来自一个失效的“最弱链”假说，它引出一个随时间单调增加的故障率函数，此函数与疲劳过程造成的不断恶化的物理表现相吻合。

公式（6.29）也可以写成 Basquin 方程形式：

$$N_i(p,S) = C_{fi}(p)S_i^{-m_i(p)} \tag{6.30}$$

其中，C_{fi}和 m_i 是随机变量，用来描述丝焊材料的疲劳行为。

因此，材料疲劳特征的不确定性包括在随机变量 C_i 和 m_i 的概率分布中。在此基础之上，材料的疲劳强度通常由一个 p-S-N 图表示，它是 S-N 曲线的同族曲线，如图 6.5 所示。此曲线表示图 6.1 中 Coffin-Manson 曲线的高循环疲劳部分。疲劳寿命的标准偏差 σ_N 随着等幅疲劳试验中应力范围的减小而增加，随着随机振幅试验中应力范围的减小而减小。为简单起见，通常把 m_i 作为一个常量考虑，因为对于某些材料种类来说，p-S-N 曲线几乎是平行的。在此情况下，我们可以认为公式（6.29）中的 C_i 是一个服从 Weibull 分布或对数正态分布的随机变量。

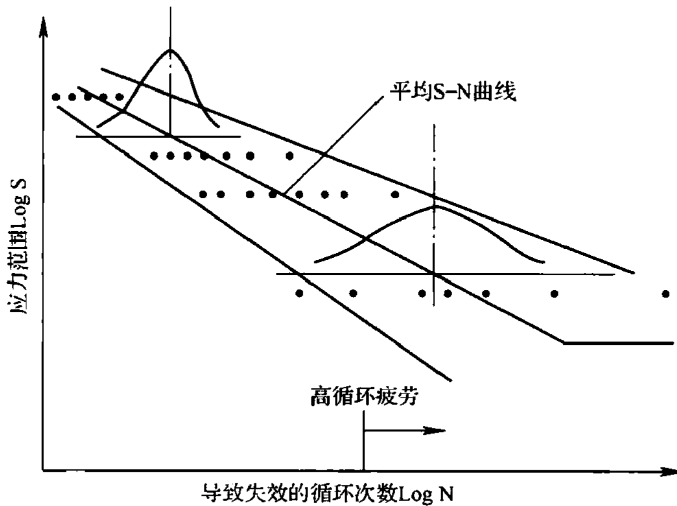


图 6.5　S-N 曲线的同族曲线

由温度循环引发的第 6.6.1 节中关键失效位置的应力范围 S_i，可以根据每一种失效机理描述为

$$S_i = H_i \Delta T \tag{6.31}$$

其中，H_i 是第 6.6.1 节中失效模型中 ΔT 系数的符号化表示。

根据公式（6.30），公式（6.29）也可以写为

$$N_i = C_i H_i^{-m_i} (\Delta T)^{-n_i} \tag{6.32}$$

其中，H_i 是其他基本变量，如材料和机械的热组成特性、丝焊的几何特征的函数。如果这些变量都被建模为一个随机变量，那么 H_i 是一个随机变量。实际上，H_i 反映了温度每变化一度的应力变化情况。

为了准确预测丝焊的寿命和可靠性，几何参数如 r、D 和 L/D 的度量以及材料特征如 E_w、E_p、E_s、α_w、α_p 和 α_s 的统计分析都需要用来确定 H_t 的分布，此过程建立在多随机变量函数的平均值和标准偏差的公式化表示之上。实际中，此方法的局限性在于它很难获得精确的标准偏差，实验研究需要把重点放在解决此问题上。

一般来说，我们还要把 ΔT 建模为一个随机变量，因为在很多情况下，使用过程中的温度循环变化在本质上是不同的。温度循环的不确定性来自于环境的变化和用服务历史描述的使用环境的变化。从用户的角度出发，环境因素是很重要的，因为它有可能影响到安全；从制造商的角度出发，在已知最低成本及可靠性的情况下，使用过程是非常重要的，因为在此过程中，设备要适应不同的区域条件。评估丝焊的可靠性需要考虑来自这两方面的不确定性：第一个方面的不确定性直接关系到疲劳损伤过程，第二个方面的不确定性可以通过设备用户的统计分析来计算。

在已知环境条件下，温度范围 $\Delta = \Delta(t)$ 可以作为一个随机过程对待，它由一系列可能的温度—时间历史组成。在任意时刻 t，ΔT 都是一个随机变量，任何温度—时间历史的度量都是一个样本函数。例如图 6.6 给出了一个空射武器的典型样本函数，它是应用时间的一个函数［Hu et al，1991］。其中，虚线是飞行高度，实线是温度。因为温度范围影响疲劳损伤程度，所以需要一个可用的累积疲劳损伤理论，以决定已知随机过程 $\Delta T(t)$ 的等效温度范围。通常需要在服务条件下进行现场数据采集，以决定分布、统计参数和其他随机过程 $\Delta T(t)$ 的特征。文献［Hu et al，1991］中使用一个三段分解法解决了此问题：

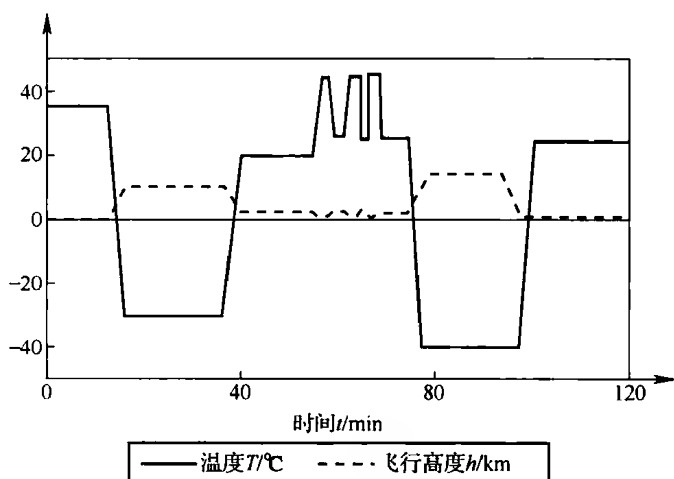


图 6.6　一个空射武器的典型样本函数（它是任务时间的函数）

$$T = u_y + Y_c\cos(2f_y t) + Z(t) \tag{6.33}$$

式中　u_y——一年的平均温度；

Y_c——季节循环的振幅，它是服从 Raleigh 分布的一个随机变量；

f_y——一年中波的频率；

t——时间参数；

$Z(t)$ ——由反应温度变化的功率谱描述。

$$f(\Delta T)=\frac{\beta}{\eta}\left(\frac{\Delta T-\gamma}{\eta}\right)^{\beta-1}\exp\left\{-\left(\frac{\Delta T-\gamma}{\eta}\right)^{\beta}\right\} \tag{6.34}$$

其中，$\beta=2.7$，$\eta=109$，$\gamma=0$。为了描绘由不同设备用户引起的不确定性，我们需要一个基于市场预测的统计数据分析。基于一些样本数据，文献［Hu et al，1991］认为样本服从 Weibull 分布。为了方便起见，假定数据服从 Gaussian 分布：

$$f(\Delta T)=\frac{1}{95.6}\exp\left\{-\left(\frac{\Delta t-96\gamma}{76}\right)^{2}\right\} \tag{6.35}$$

其中，ΔT 的平均值为96°C，标准差为38°C。如果不同用户需要的设备数量从市场预测中估计，每个用户的一般温度范围就可以得到了。由于第二种类型的不确定性存在，ΔT 的分布可以通过拟合优度试验方法来估算。

丝焊热疲劳的不确定性来自于丝焊的材料行为、装配尺寸和温度变化。为了精确预测疲劳寿命并分析可靠性，必须要考虑这些不确定因素。

6.6.3　疲劳寿命和可靠性预测

公式（6.32）把丝焊中第 i 种失效机理的疲劳寿命描述为一个四参数随机函数：

$$N_i=C_i(H_i\Delta T)^{-m_i} \tag{6.36}$$

其中，H_i 和 ΔT 统计独立于其他两个变量，但 C_i 和 m_i 却是互相关联的，因为它们来自同样的疲劳试验数据。

为了便于描述，我们假定 ΔT 拥有一个指定值，μ_i 为常数，使用多随机变量理论（参见本书的第2章、第3章），我们就可以得到：

$$\mu_{N_{ai}}=\mu_{C_i}(\mu_{H_i}\Delta T)^{-m_i} \tag{6.37}$$

$$\sigma_{N_{ai}}^{2}=\left(\frac{\mu_{N_{ai}}}{\mu_{C_i}}\right)\sigma_{C_i}^{2}+m_i\left(\frac{\mu_{N_{ai}}}{\mu_{H_i}}\right)\sigma_{H_i}^{2} \tag{6.38}$$

因此，当其余参数的平均值和标准偏差已知时，我们就可以获得预期寿命 N_i 的平均值 $\mu_{N_{ai}}$，和标准偏差 $\sigma_{N_{ai}}$。

因为随机变量 H_i 和 ΔT 是统计独立的，所以，N 的累积分布函数（Cumulative Distribution Function，CDF）为

$$F_N(N)=P\{N<n\}=\iint p(C_i)p(v)\mathrm{d}C_i\mathrm{d}H_i \tag{6.39}$$

明确的积分表达只能从 C_i 和 H_i 的一些特殊分布中获取。

通常，循环温度范围的振幅平均值是时间的函数。这个温度振幅变量对丝焊疲劳的影响可以通过累积损伤规律计算。此规律把复杂载荷历史下的疲劳行为和恒幅载荷下的已知统计行为关联了起来。线性损伤累积模型，如 Miner 法则被广泛采用，但这些模型没有考虑应力范围在疲劳寿命上的次序效应。Miner 法则在变幅应力下预测未来疲劳损伤。参照第6.4节中的标记方法，把 DM 用作损伤参数，我们可以得到：

$$DM=\sum_{j=1}^{k_b}\Delta D_j=\sum_{j=1}^{k_b}n_j/N_j \tag{6.40}$$

式中　ΔD_j——由于第 j 个恒定应力块或应变范围 S_j 引起的损伤增量；

n_j——第 j 个块内的总循环数；

N_j——在应力循环 S_j 下，由第 i 种机理［为了方便起见，在公式（6.40）中省略了 i］引起失效的循环数；

k_b——总块数。

当损伤累积到超过容限——也就是说，当 $DM>1$ 时，失效就会出现，把公式（6.36）代入公式（6.40），可得：

$$DM=\frac{H_i^{m_i}}{C_i}\sum_{j=1}^{k_b} n_j \Delta T_j^{m_i} \tag{6.41}$$

令 $n_j=N_i p_j$（p_j 是温度循环 ΔT_j 出现的相对可能性），令公式（6.40）中的 $DM=1$，那么，公式（6.41）就变为

$$N_i=\frac{C_i H_i^{-m_i}}{\sum_{j=1}^{k_b} p_j \Delta T_j^{m_i}} \tag{6.42}$$

把第 i 种失效模式的等效温度循环定义为

$$\Delta T_{eq,i}=\left\{\sum_{j=1}^{k_b} p_j \Delta T_j^{m_i}\right\}^{\frac{1}{m_i}} \tag{6.43}$$

公式（6.43）与公式（6.36）的形式相同：

$$N_i=C_i(H_i \Delta T_{eq,i})^{-m_i} \tag{6.44}$$

如果把温度范围定义为一个连续随机变量，那么等效温度范围为

$$\Delta T_{eq,i}=\left\{\int_0^{\Delta T_{max}} p(\Delta T)\Delta T_j^{m} \mathrm{d}(\Delta T)\right\}^{\frac{1}{m_i}} \tag{6.45}$$

其中，$p(\Delta T)$ 是 ΔT 的概率密度函数（pdf）。

把公式（6.43）和公式（6.45）中的 m_i 作为一个已知其值的确定数量考虑时，$\Delta T_{eq,i}$是第 i 种失效机理的常数，$(\Delta T_{eq,i})^{m_i}$是 ΔT 的第 m_i 个时间片段。因此，变幅温度循环问题就变成了一个恒温振幅热疲劳问题，然后就可以利用公式（6.37）和公式（6.38）进行计算了。尽管如此，从制造商的观点出发，等效温度循环$\Delta T_{eq,i}$仍然不是一个常量，而是一个随机变量，计算它的分布需要分析用户以往的统计数据。因此，在公式（6.44）中，即便是把 m_i 作为确定数量考虑，仍然存在三个随机变量。此时，疲劳寿命的平均值和方差表示为

$$\mu_{N_i}=\mu_{C_i}(\mu_{H_i}\mu_{\Delta T_{eq,i}})^{-m_i} \tag{6.46}$$

$$\sigma_N^2=\left(\frac{\mu_{N_i}}{\mu_{C_i}}\right)^2\sigma_{C_i}^2+m_i\left(\frac{\mu_{N_i}}{\mu_H}\right)\sigma_{H_i}^2+m_i\left(\frac{\mu_{N_i}}{\mu_{\Delta T_{eq,i}}}\right)\sigma_{\Delta T_{eq,i}}^2 \tag{6.47}$$

一般来讲，用封闭形式方程来表示疲劳寿命的分布是非常困难的，但是，通过使用计算机模拟可以计算它的数值，例如 Monte-Carlo 方法。只有在一些特定情况下，才能给出寿命分布的显示表达式。例如 C_i、H_i 和 ΔT_i 就完全可以用一个对数正态分布模拟，即

$$\mu_{N_i'}=\mu_{C_i'}-m_i\mu_{h_i'}-m_i\mu_{t_i'} \tag{6.48}$$

其中，$N_i' = \text{Log}N_t$，$C_i' = \text{Log}C_i$，$h_i' = \text{Log}H_i$，$t_i' = \text{Log}\Delta T_i$，相应的标准偏差为

$$\sigma_{N_i'} = (\sigma_{C_i'}^2 + m_i^2\sigma_{h_i'}^2 + m_i^2\sigma_{t_i'}^2)^{\frac{1}{2}} \tag{6.49}$$

无论是用封闭方程计算还是用每一种失效机理的仿真法确定疲劳寿命，只要确定了运行寿命 P（N_t）的分布概率，就可以为每一种机理绘制可靠性相对于疲劳寿命的曲线了。此曲线可以确定占主导地位的失效机理，还可以估计在所要求的可靠性级别下的疲劳寿命。

可靠度 R_t（N）被定义为丝焊在某数量运行热循环内可用的概率，每种失效机理的 N 为

$$R_i(N) = \int_N^{\infty} p_i(N)\mathrm{d}N \quad (i = 1,2,\cdots,5) \tag{6.50}$$

在失效机理以串联形式出现且统计独立的情况下，丝焊的可靠度可计算为

$$R(N) = \prod_{i=1}^{5} R_i(N) \tag{6.51}$$

然而，这些失效机理中的一些可能是相互关联的，例如 C_1、C_2、m_1 和 m_5 都来自相同的试验数据组。对于相关联的失效机理，丝焊组装的可靠度可以估算为

$$\prod_{i=1}^{5} R_i(N) \leqslant R(N) \leqslant \min_{i=1}^{5} R_i(N) \tag{6.52}$$

这些公式的数值应用参见文献［Hu et al，1991］。文献作者 Hu et al 为样本假设了一些输入数据，根据计算结果绘制的可靠度曲线样本如图 6.7 所示。

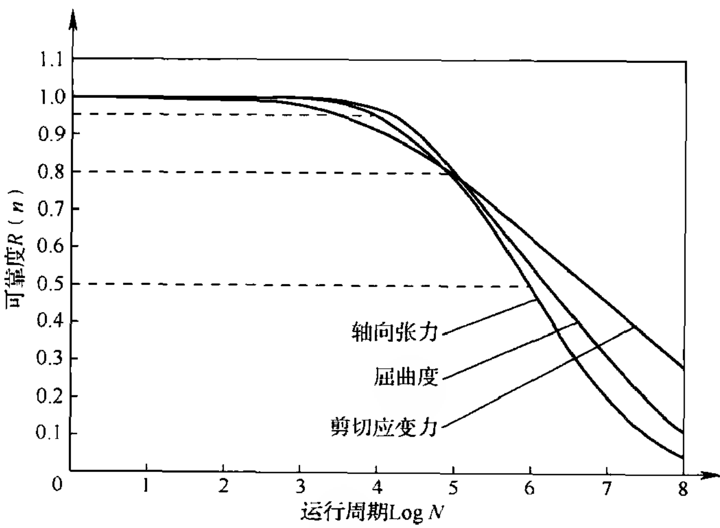


图 6.7　可靠度曲线样本

6.7　鉴定试验和加速试验

我们在第 6.2 节中讨论了损伤模型，从这些模型中获得的可靠性预测接近于最佳

值，其准确性取决于输入参数及其不确定性数据库的精度。因此，最重要的是在原型开发阶段，通过广泛的鉴定试验和重设计对预测进行验证。在设计或制造规范发生改变时，这种验证都要重复进行。鉴定试验的目的是在实际生命周期载荷下验证产品是否真正地达到了预期的可靠性。换句话说，鉴定试验试图找出产品持续存活超过延长期（通常是产品的设计寿命）的概率。因此，鉴定试验审查设计规格满足可靠性目标的能力。通常，这项工作在加速应力下完成，以满足压缩试验时间的要求。一个设计良好的可靠性鉴定程序可以节约财力，并可以在新产品开发时或成熟产品面临制造和工艺改变时快速周转资金。

在拟使用环境中运行时，很多现在的工程硬件都具有很高的可靠性。为需要更长寿命的产品调查磨损失效机理，度量其可靠性或许会成为一项挑战。因为在实际运行条件下，需要用较长时间的试验周期来获取充分的数据，才能决定实际失效特征。如何为高可靠性设备取得有意义的质量试验数据是个问题，解决办法之一是加速磨损试验，有时被称为加速试验或加速应力寿命试验。然而，对过应力机理进行可靠性鉴定时，预计过应力载荷的单个循环可能就足够了，并且也不需要对试验数据加速。

为磨损失效机理进行的加速试验，涉及到把载荷和应力加速到比正常运行条件更苛刻时衡量试验设备的表现，以在短时间内引发失效。在以下条件中，此类试验的目的是加速与时间相关的磨损失效机理以及损伤累积，以减少失效时间：

① 加速环境中的失效机理和失效模式，使其与使用环境中观察到的一样（或有数量上的关系）。

② 在具有一定的把握时，可以从加速环境量化推测出使用条件。

③ 材料的工程特性在加速应力下表现良好。

④ 失效在运行级别和加速试验水平中的分布相似。

科学的加速试验方法首先要确定相关的磨损失效机理。选择直接引起时间关联失效（Time-Dependent Failure）的应力参数，把它作为加速参数。应力参数通常被称作加速应力。常见的加速应力包括：热应力，如温度、温度循环和温度变化率；化学应力，如湿度、腐蚀、酸和盐；电应力，如电压、电流和功率；机械应力，如振动载荷、机械应力循环、应变循环和冲击等。

加速环境可能包括这些应力中的某一种应力或是这些应力的某一种组合形式。为了说明应力组合的结果，需要对它们之间的相互关系以及每一种应力对整体损伤的影响进行非常清晰、量化的了解。

加速试验的技术涉及到选择失效机理以及适当的加速应力；确定试验程序和应力级别；确定试验方法，如采用恒定应力加速，还是步进应力加速；进行试验；说明试验数据，包括把加速试验结果推测到正常运行条件。因为失效模式和材料特征存在不确定因素，量化推测通常较困难。然而，试验结果却通常能为设计师提供好的量化失效信息，使得设计师通过改变设计和（或）工艺来改进硬件产品的可靠性。

一些加速参数可以引发特定机理导致的失效。例如温度和湿度都可以加速腐蚀，机械应力和温度都可以加速蠕变。此外，单独的加速应力可以同时引发一些磨损机理导致

的失效。例如温度可以加速损伤累计，其原因不仅是电迁移，也包括腐蚀、蠕变等等。正常运行条件下，占主导地位的失效机理可能会因为应力的增加而失去其主导地位；相反，在正常使用条件下潜在的失效机理，可能会在加速条件下引发设备失效。因此，要仔细计划加速试验，以用它来表现实际使用环境和运行条件，但它不能引起多余的失效机理或非表现性的物理和材料行为。

通常，应力的加速程度由一个加速因子控制，它被定义为正常使用条件下的寿命与加速条件下寿命的比值。要为有问题的硬件专门制定加速因子，从一个加速变换中估计其值，此加速变换为所有的硬件参数定义了加速应力和寿命减少之间的关系函数。很明显，变换函数需要建立在第6.2节中所讨论类型的量化失效模型的基础之上。

在鉴定和验证计划中，对失效样本进行详细的失效分析是一个关键性步骤。没有这些分析及其对设计师修复行为的反馈，鉴定计划的目标就无法实现。换句话说，仅收集统计的失效数据是不够的。单独的统计数据不能提供相关失效机理的信息，也不能提供阻止这些失效机理的方法。关键是使用试验结果设计出更健壮和更具成本效益的产品。

6.8　降额和后勤决策

鉴定试验的反馈结果需要用到硬件的重新设计中去，以便获得更好的可靠性。在成熟的技术中，这类可靠性增长是产品开发的一个内在组成部分。其过程是迭代进行的，一直要延续到可靠性和（或）成本效益目标已实现。当产品不可能进行更好的改进时，如果可靠性仍然不能满足需求，那么，降额和设计冗余或许是改进产品可靠性的下一个选择。

降额（De-Rating）是一项要么相对于额定强度减小设备或结构的运行应力，要么相对于所分配运行应力水平增加强度的技术。它把硬件运行载荷的上限设定在额定承受能力之下，这样做可以减小应力。例如电子硬件的制造商通常会设定供给电压、输出电流、功率消耗、结点温度和频率的限制。设备的设计师选择一个可选的零部件或是做出设计改变，以保证特定参数的运行条件（如温度）一直处于额定水平之下。这样，我们就可以说“已经对零部件的热应力进行了降额”。

降额因子通常定义为已知应力参数的额定水平与实际运行水平的比值，它实际上是一个安全裕度或容差裕度。容差裕度由所有可能发生的失效的危急度决定，它由可靠性模型本身和输入到其中的内容所固有的不确定性度量。理想状况下，要保持此裕度为最小值，以维持设计的成本效应。可靠性工程师需要对此负责，以尽可能清晰地确定额定强度、相关运行应力和可靠度。

为了提高有效性，降额标准必须以正确的引发失效机理的应力参数作为目标。此外，失效模型和实验室试验的输入内容用来衡量这些参数的精确度，必须对这些输入内容进行定义。允许的运行应力可以通过相关失效机理的量化建模和失效物理概念与设计强度关联起来。可能还需要将现场测量和建模仿真结合起来确定失效位置的实际运行应力。一旦量化了失效模型，就可以确定降额对已知载荷下零部件的有效可靠性的影响。

降额和可靠性之间的量化关系可以让设计师和用户有效地调整安全裕度，以适应零部件的危急度，这样可以使零部件的功能性能力（Functional Capacity）的利用更具成本效益。

可靠性预测的价值远比开发设计模型的价值大得多。很多生命周期的后勤决策都是由可靠性预测驱动的，其他还包括成本、可修理性和预防性替换日程表等。可靠性影响后勤功能的例子包括可靠性分配、维修行为日程安排、备件设计、保修政策和作废计划等。

可靠性物理预测的其他后勤限制了产品开发的时间。详尽的建模和对全新技术的质量检测都需要竞争市场中不可能存在的时间和资源。在此情况下，通常会为新产品引入一个不成熟的半开发阶段。此时，继续进行鉴定测试是很有必要的，这样才能促使可靠性成长。随着工程人员越来越多地采用计算机辅助建模、自动化预测工具、改进的材料性能数据库，这些工具的应用将使得可靠性预测变得更简单。对于相关失效机理的合理理解、广泛的材料试验程序在实现这些高级建模能力方面起到了关键性作用。

6.9　制造问题

制造、工艺和装配在很大程度上影响着硬件的质量和可靠性。不恰当的装配和制造技术会导致产品的缺陷、瑕疵和残余应力。在零部件寿命的晚期，残余应力表现为潜在失效位置或“应力集中区”。在装配和制造过程中，缺陷和应力在为了确定它们所需的运行中会影响硬件的可靠性；在设计和开发阶段，确定这些缺陷和应力可以帮助设计分析师提前考虑它们。审查制造工艺的特点涉及到两个关键步骤：第一，和在设计鉴定中一样，对工艺进行鉴定，以保证制造规范没有过分影响硬件的长期可靠性。第二，进行批次筛查，以保证所有制造相关参数的变化都在所定义容限内。也就是说，筛查保证了产品的质量，它在潜在缺陷出现之前使其暴露，从而改进了硬件产品的短期可靠性。

6.9.1　工艺鉴定

像设计鉴定一样，此鉴定程序可以在原型开发阶段进行，目的是确保标称制造规范和容限能使硬件产生可接受的可靠性。一旦通过了工艺鉴定，只有当工艺参数、材料、制造规范或人力因素改变，才需要对工艺进行再鉴定。

工艺鉴定试验应该与设计鉴定试验中的加速磨损试验有相同设置。在设计鉴定试验的过程中，过应力试验可用来在预期过应力载荷下对产品进行鉴定。另外，还可以利用过应力试验来确保制造工艺没有降低硬件的固有材料强度。然而，这些试验的作用是补充而不是取代加速磨损试验程序，除非已经明确得知过应力试验结果和磨损现场失效数据的物理关系。

在设计鉴定中，现场失效机理的加速是主要目标，所有在第 6.4 节中给出的警告都可以应用到这里。试验结果对于失效分析和闭环修复行为是最重要的，它们也最有效地利用了试验结果。

6.9.2　工艺性、工艺变化和缺陷、产出

对制造缺陷的控制和校正是生产和过程控制工程师所关注的，而不是设计师所要关

注的重要内容。然而，在并行产品开发的理念和环境中，硬件设计师必须了解材料限制、可用工艺和制造工艺能力，并以此来选择材料、构建改进可生产性的结构，帮助减少缺陷发生，增加产出、提高质量。

因此，没有一个明确的关于制造缺陷和可接受约束的讨论，就不能完成生产规范。可接受质量的门槛是什么？什么样的产品是不合格的？对于这些问题，可靠性工程师必须有的明确定义。此处，我们认为导致硬件表现和可靠性有妥协表现的不合格是一种缺陷。失效机理模型提供了一个方便的工具，可以用它来制定这些标准。对可靠性分析师来说，重要的是从规范中找出那些折中性能和可靠性的偏差以及那些良性的可接受的偏差。此处的重点是在另外的鉴定过程中，由于过度的批次变化而产生的不良质量和缺陷。这些变化通常是由不好的过程控制产生的。

任何随时会损害或有可能损害产品功能的工艺过程（制造或装配）都会导致缺陷出现。缺陷可能出现于单个工艺过程中，或者表现为一系列工艺过程的结果。过程的产出是产品的零部件，在后续的制造程序或产品生命周期的过程中，这些产出必须是可使用的。工艺的累积产出由每个工艺步骤的单个产出结果的总和决定。有时候，缺陷的根源不是很明显，因为直到产品到达下游某个工艺步骤，工艺过程导致的缺陷都可能不会被发现；如果没有对产品进行筛选，结果更是如此。

当某个过程在控制（在规范中运行）之中时，观测缺陷率与过程工程师的预测将会相同。如果工艺有任何部分不在控制之中，产品就会产生缺陷，我们必须要理解工艺步骤和缺陷的关系。好的过程控制能优化产出和可靠性。然而，即便获得了高产出，也不一定能保证产品的可靠性。因此，初步验收标准和可靠性之间就可能存在不匹配的现象。工程师必须在设计可靠性的同时，必须使工艺保持在控制之中。

通常，简化制造和装配过程，以减少工艺缺陷是可行的。然而，当工艺变得更加复杂时，就需要对过程进行检测和控制，以保证产出无缺陷产品。判定工艺是否在容限范围内的界限通常被称为工艺窗口（Process Window），它是按照工艺中所要控制的独立变量以及工艺对产品的影响或相关的产品变量定义的。

我们的目标是了解每个工艺变量对所有产品参数的影响，以此来为工艺制定控制范围——也就是说，其中的缺陷率已拥有引发失效潜力的一些变量范围。在定义工艺窗口时，要使每个工艺变量的上、下限超过那些已确定的、能引起缺陷的值。通过缺陷试验和缺陷原因分析，使制造过程包含于工艺窗口中，利用过程控制（如闭环修复系统）来清除缺陷。建立一个有效的反馈途径是至关重要的，它可以用来报告过程相关的缺陷数据。一旦完成了这项工作，也就决定了工艺窗口，它将成为一个面向过程操作员的反馈系统。

一些相互影响的工艺参数产生的缺陷，不同于这些参数独立作用产生的缺陷。这种复杂的情况要求用一个试验矩阵来计算各种工艺参数的相互作用。某些情况下，直到工艺步骤的后面部分才能发现缺陷。因此，缺陷能引起产品的驳回、返工或失效，此时的产品已附加了大量价值。这些由于缺陷产生的成本会减少产出，它附加到隐藏的工厂成本中也会减少投资收益。我们需要特别关注所有关键的工艺，这样才能通过工艺控制来

消除缺陷。

制造和装配质量的策略取决于对每个单独工艺步骤的理解和控制以及策略对产品的影响。我们的目标是减少缺陷发生的概率，使工艺步骤的监测更加容易，改进硬件的可制造性。

6.9.3 工艺验证试验和统计过程控制

工艺验证试验通常被称为筛选。筛选要求对所有制造产品 100% 的审查，以发现或暴露缺陷，它要求合格产品有批次生产基础。筛选的目标是在产品产出之前，预先发现潜在的质量问题。原则上讲，完全处于生产控制之下的过程是不需要它的，然而由于控制程序中的不确定因素，筛选通常被当作安全网使用。如前所述，现场失效机理的量化模型完全可以为缺陷定义合格界限。

一些产品的失效表现为一个多峰概率密度函数（Multimodal Probability Density Function），它的第二个峰值位于产品服务寿命的早期阶段，其原因是问题材料的使用、制造和装配技术的不良控制或不正确运转。这种现象通常被称作早期失效（Infant Mortality）。适当地应用筛选技术，可以成功地发现或暴露这些失效，并可以消除或减小它们在使用场合出现的概率。如果失效概率密度函数中没有一个主要峰值，那么筛选或许是一个多余的成本消耗项。很难对由于突发事件，如天灾（闪电、地震等）引起的失效进行设计或筛选，以提高成本效益。只有在生产的早期，才考虑筛选；如果需要的话，仅在产品将出现早期失效时才考虑使用。在产品缺乏健壮设计、成熟产品面对新指定的零部件、材料或工艺时，使用筛选是比较合适的。

因为筛选是在 100% 零件的基础上进行的，所以开发无害于完好零件的“筛网”就显得尤为重要。应力筛选涉及到产品的运行应力，它有可能在额定运行范围之外。因此，最好的筛网应该是无损检测技术，如显微视觉检查、X 射线、声波扫描、C 扫描、核磁共振、电子顺磁共振等等。如果应力筛选是不可避免的，过应力试验就优于加速磨损试验，因为后者有可能会消耗好零件的使用寿命。

应力筛选不一定非要模拟失效的现场环境，它甚至可以利用现场条件中的缺陷所引发失效机理的相似机理。相反，筛网需要利用最方便、最有效的失效机理，以便在现场模拟早期缺陷。很明显，这要求我们能觉察到硬件中可能出现的缺陷，还要熟悉相关的失效机理。

在应力筛选中，如果对好零件的损伤是不可避免的，那么就必须基于失效机理模型对筛选损伤进行量化估计，这可以让设计师说明使用寿命在此过程中的损失。筛选过程中的应力级别一定要适用于特定的硬件产品。在鉴定试验中，可以根据失效机理量化模型的变化决定要使用的筛选参数。

与鉴定试验不同的是，如果在确认操作将引发缺陷后立即执行筛选，它将拥有最大的有效性。鉴定试验倾向于在产品完成或十分接近于最后运行时进行；另外，如果只在最后阶段进行筛选，也就是当所有运行都已完成时进行筛选，它的有效性最低，因为失效分析、缺陷诊断和故障维修都很困难，并且还会损害修复行为。此外，如果在制造过程早期发现了缺陷，后续的通过新材料和工艺附加的价值都会被浪费掉，这也会额外地

增加运行成本，降低生产效率。

不可否认，这种方法也存在一些缺陷。每一个制造基地处的筛选成本都是很昂贵的，小批量作业尤其如此。另外，零件会随着制造步骤经历重复的筛选载荷，这样就会增加完好零件磨损破坏累积的风险。为了得到筛选矩阵，需要尽可能多地描述每个筛选试验可用的缺陷和失效机理，最佳状况应该是在成本效益分析、风险和关键性缺陷中搜索这些缺陷和失效机理。所有缺陷都要追溯到引发变异的根源。

应力筛选为资本、运行开销和循环时间带来了大量的负作用，它所能带来的益处会随着产品接近于成熟而减小。任何将要进行的应力筛选都必须有资金和人力来为所有失效单元确定根本原因，并采取合适的修复行为。应该由设计、制造和质量小组来确定应力筛选类型。虽然，生产的早期有可能需要进行应力筛选，但我们还是强烈建议在失效修复过程中制定一个在早期减小样本数量的筛选计划。

如果几乎所有的产品都在一个适当设计的筛选试验中不合格，那么设计就可能是错误的；如果有很多产品不合格，就需要对制造过程进行修正；如果只有少量产品不合格，那么工艺就可能在容限范围内，观测到的故障量或许会超过对其进行修复所需的设计和生产工艺资源。在工艺成熟、筛选拒绝降低标准时，需要由经济因素来驱动筛选决策，用取样程序代替筛选可能是比较合适的选择。

在鉴定试验中，失效分析、适当的反馈以及合适的修复行为都是非常必要的，它们用来确保及时清除导致缺陷的原因。然而，对于合格的产品来说，只有改进批次间的变化、产品质量和短期可靠性时才使用筛选。只有通过有前瞻性的设计和工艺改变，才能改进产品的长期可靠性，并且这只能在开发过程的早期阶段，通过及时的鉴定试验程序来完成。在产品生命线早期，持续的设计和工艺改变的成本非常大，制造商并不想付出这些成本。

6.10 总结

只有优秀的产品设计、容限内良好的工艺能力以及来自供应商合格的零部件和材料才能保证产品的高度可靠性，同样，供应商也要有良好的、在容限范围内的工艺能力。为高可靠性建立有效的设计、工艺规范和容限，所有相关失效机理的量化认知和概率模型为此提供了一个方便的工具。精确的可靠性预测要求精确的应力分析，并要有为不同使用类型的预期生命周期载荷建立的精确数据库。在制造、加速试验和使用过程中，预期的所有载荷范围内的材料构成和损伤特征也是必不可少的。后勤任务包括维修计划、保修定价等。对于这些任务的前期预测来说，适当的可靠性预测也会有所帮助。

科学的可靠性评价总是需要加速鉴定试验作为补充。要仔细考虑制造工艺对质量和可靠性的影响。只有使用合格的程序，谨慎并持续地将其控制在最小的变异和缺陷中，才能保证硬件产品的良好制造。筛选和鉴定试验拥有截然不同的目标，因此，所选择的筛选和试验必须要适用于特定的执行目的。一般来讲，鉴定试验一定会触发相同的失效机理，这些机理会引起现场失效，进而影响硬件产品的长期可靠性。只要筛选试验能成

功移除目标缺陷，而且不损伤好的零部件，它就可用于任何可用它来清除的失效机理。

硬件可靠性不是偶然形成的，也不是运气问题，在每个设计、开发和制造阶段，工程人员必须要付出有意识的、系统的、严格的努力才能获得。不可否认的是，这样的方法要求工程人员具有广泛的建模和材料行为知识。也可以说，没有这样系统的方法和对细节的认真关注，我们将不能持续控制复杂且昂贵的工程硬件的可靠性。

参考文献

Broek, D. 1986. Elementary engineering fracture mechanics. Boston: Martinus Nijhoff. Dasgupta, A., and H. Haslach. 1993. Mechanism design failure models for buckling. IEEE Transactions on Reliability 42.

Dasgupta, A., and J. M. Hu. 1992. Failure – mechanism model tutorials: (i) Excessive elastic deformation; (ii) Plastic deformation; (iii) Brittle fracture; (iv) Ductile fracture. IEEE Transactions on Reliability 41 (1-4): 149-154; 168-174; 328-335; 489-495.

Dumoulin, P. 1982. Metal migration outside the package during accelerated life testing. IEEE Transactions on Components, Hybrids Manufacturing Technology 479.

Engel, P. 1993. Failure models for mechanical wear modes and mechanisms. IEEE Transactions on Reliability 42: 9-16.

Ghanem, R., and P. D. Spanos. 1991. Stochastic finite element methods. New York: Springer-Verlag.

Haugen, E. B. 1980. Probabilistic mechanical design. New York: John Wiley & Sons. Hertzberg, R. W. 1989. Deformation and fracture mechanics of engineering materials. New-York: John Wiley & Sons.

Hu, J. -. 1994. Physics-of-failure based component qualification of automotive electronics. In Reliability, maintainability, and supportability, SAE.

Hu, J. -., M. Pecht, and A. Dasgupta. 1991. A probabilistic approach for predicting thermal fatigue life of wirebonding in microelectronics. ASME Journal of Electronic Packaging ii3 (3): 275.

Jones, R. M. 1975. Mechanics of composite materials. New York: McGraw Hill.

Kapur, K. C., and L. R. Lamberson. 1977. Reliability in engineering design. New York: John Wiley & Sons. Lewis, E. E. 1987. Introduction to reliability engineering. New York: John Wiley & Sons.

第7章 软件可靠性

7.1 引言

软件有着非常重要的作用，它已经成为我们生活中不可或缺的部分。因为软件所涉及的领域和类型都很众多，软件可靠性已经逐渐成为一个关键问题。

对于很多严重依赖于软件的产品，安全是关键问题，这些产品包括：飞行系统、航空交通管制系统、核电站中用于帮助操作员诊断事故根源并确定缓解行为的产品、卫星遥控器、医疗产品等等。即便在不影响安全的时候，软件失效也可能会产生严重后果，如信息系统中有价值数据的丢失、管理不善的银行交易以及记账错误。现在，软件开发公司要对他们的产品质量负责，产品开发成本也越来越受到软件成本的驱动。有趣的是，虽然硬件在低成本下很可靠，但在很多应用领域中，软件却正在取代硬件。

所以，在这样的情况下，已有很多技术可以对软件可靠性进行定性、定量地度量和改进。这些技术的目的在于改进软件产品的质量，同时改进软件开发流程的质量，包括：用于设计更好的代码的软件工程；有效移除故障的测试技术；为细化软件能力规范，并保证最终产品满足需求的形式化方法；表示产品复杂度、度量软件开发流程状况的定性指标以及评价可靠性的量化模型。这些技术都对软件可靠性有所贡献，本章将介绍这些内容。

本章首先给出软件、软件可靠性、软件质量和软件安全的定义，描述软件开发生命周期过程，并一针见血地指出在过程中每个阶段发生的软件错误的机理。然后再阐述一些决策技术，这些技术可以改进给定软件产品的可靠性，最后介绍了一些评价软件可靠性的定性量度和量化模型。

7.2 相关定义

软件（Software）：在电子及电气工程师协会（The Institute of Electrical and Electronics Engineers，IEEE）的标准软件工程术语表［Standard Glossary of Software Engineering Terminology，1983］中，软件的定义是计算机程序、过程和规则，它可能与文件编制以及与电脑产品运行相关的数据有关。

从定义可以明确地看出，软件不仅仅是运行程序的几行代码。也可以说，软件是除了运行软件的物理硬件之外的所有东西。

固件是软件的特殊形式。根据［IEEE，1983］，固件是：

① 计算机程序以及加载在某种存储器中的数据，在运行中，这些数据不能由计算

机动态地修改。

② 硬件，它包含了一个计算机程序以及一些不能在用户环境改变的数据；包括在固件中的计算机程序和数据被归类为软件（包含计算机程序的电路）和硬件数据。

③ 存储在一个只读存储器中的程序指令。

④ 一个组件，它包括一个硬件单元和一个计算机程序，两者结合起来形成一个功能实体，在正常运行时，不能改变它的配置。计算机程序以集成电路的形式存储在硬件单元中，它有固定逻辑配置，此配置能满足特定的应用或运行需求。

因此，软件包括嵌入在系统中的微编码或微程序。开发或制造固件的项目（或许是不可计算的，例如查找表格），也可能会受到软件可靠性行为的影响。

软件可靠性（Software Reliability）：人们已接受的软件可靠性定义[⊖]与硬件可靠性相似。根据根据［IEEE，1983］，软件可靠性是：

① 在特定条件下、特定时间内，软件将不会引发产品失效的概率。此概率是产品的输入内容以及使用状况的函数，还是当前软件中存在故障的函数。输入到产品的内容决定了是产品否存在故障，如果有，使用者就会遇到它。

② 在特定条件下、特定的时间段内，程序实现所要求功能的能力。

现在已经出现，将来也还会继续出现一些关于软件可靠性定义的争论，出现这些争论的主要原因是人们已经选择把时间作为度量可靠性的基础。没有争议的是，“持续时间”与应用程序相关，例如需要在长时间内实现其功能的操作系统，但是，它或许不适用于诸如编译程序和科学应用之类的内容。

让我们假设：一个科学应用程序的代码由两个不同的开发小组完成。他们的产品很可能将成为两个不同的软件：S_1 和 S_2。我们把失效定义为不正确的输出。假如两个软件的输入是相同的，S_1 在 T_1 中运行，S_2 在 T_2 中运行，$T_1 < T_2$。如果两个输出 O_1 和 O_2 都不正确，根据 IEEE 对软件可靠性的定义，可以推测出一个毫无意义的结论：软件 S_1 比 S_2 不可靠。因此，对于特定的应用程序，一个合理的可靠性定义是：在特定数量的运行次数中，软件能完成任务的概率。

从这两个含义中可以清晰地看出：减少或消除由于软件产生的产品失效，减少并避免软件中的故障可以实现软件可靠性。我们在［IEEE，1983］中发现了一个专有词汇，它把故障、失效与软件开发相关活动联系了起来：

① 错误是人的行为，它导致软件包含故障，例如对用户需求的省略和误解，设计资料中对需求的省略或不正确转换，或是一行代码、数据表格或分支条件的不适当编码。

② 在软件中，故障是错误的表象；如果遭遇一个故障，它有可能会引发失效——“bug”的同义词。

③ 失效是指产品或部件在特定约束下不能完成所要求的功能，用户在测试或实际

⊖ 这个软件可靠性定义被称为“软件可靠性的用户定义”，它与“软件可靠性的开发者定义”相对应，后者常用于“每千行源代码中错误的数量”这一概念。

使用中会观察到失效。

虽然“缺陷”一词在文献中广泛使用，但是除了向读者介绍“故障”时，IEEE 没有对它进行详细说明。其他作者可能会区分缺陷和故障，但本章将遵循 IEEE 术语，除非在特定资料来源中使用“缺陷”一词。

在变成一个可运行的产品之前，软件会出现错误和故障，所以要在此之前规划并启动软件可靠性项目。软件可靠性项目有五重目标：

① 在软件到达测试、运行使用阶段之前，阻止错误和故障的发生。

② 使用在软件开发中得到的信息消除未观察到的错误和故障。

③ 减轻软件失效的影响，特别是严重性后果。

④ 收集项目数据，更好地理解允许错误和故障发生的条件和现象，并使用这些经验改进软件开发流程。

⑤ 对交付代码进行可靠性估计。

在一个缺乏有效的可靠性计划的软件开发工作中，错误的级联方式如图 7.1 所示。开发工作中，在确定需求时发生的错误会在代码中变成失效；在设计和编码阶段发生的错误最终也会变成代码故障。因为在测试中（它本身也将包括错误，例如不适当的测试，或某些区域测试的失效），所有故障都是潜在的失效，测试着重于找到并改正那些由于故障级联引发的失效，或许这些故障已在早期阶段确定并消除。尽管如此，在测试阶段出现的故障仍可能相互遮掩，这进一步增加了把这些故障带入到运行代码中的风险。很少有项目会拥有测试资源或列出所有级联错误可能造成影响的计划。

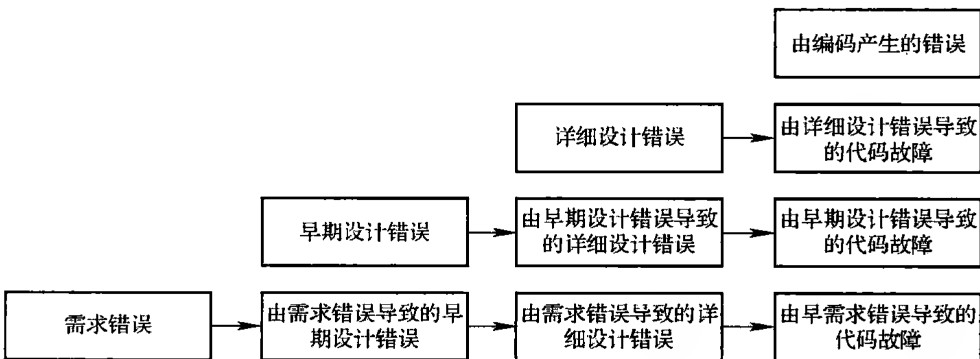


图 7.1　开发工作中错误的级联

在（临近）错误或故障发生的阶段，有效的软件可靠性计划能确定并改正错误和故障。这种方法能最小化级联错误的影响，从而使测试工作可以专注于开发早期不容易解决的产品界面和整合问题。

软件质量（Software Quality）：根据［IEEE，1983］，软件质量的定义为“软件产品满足给定需求的所有特征和功能部件的总和，例如符合规范要求”。除了可靠性，美国国防部标准部 2167、防御系统软件开发 1985（DoD1985）还描述了其他适用于软件产品 10 个质量因素。这些因素的定义在表 7.1 列出。

表 7.1 DOD-STD-2167 中确定的质量因素

质量因素	定 义
正确性	软件没有设计缺陷、代码缺陷的程度——即无故障程度
效 率	软件以最小的计算资源消耗运行既定功能的情况
灵活性	改进软件的难易程度
完整性	软件阻止或控制无授权访问，并通知数据或代码的情况
互用性	两个或更多产品交换信息的能力
维护性	软件维护的难易程度
便携性	软件从一个计算机系统或环境转移到另一个的难易程度
可复用性	一个模块在多个应用程序中的使用情况
可测试性	软件测试的难易程度
可用性	软件符合人类工程学的程度

虽然不同的组织和计划所要实现的可靠性目标是不一样的，但在实现软件可靠性过程中，软件质量保证（Software Quality Assurance，SQA）组织起到了明显的作用。在小的企业组织中，可能由软件开发人员展开 SQA 活动；而在大型企业组织中，一个项目组的 SQA 组织负有为项目制订计划并开展 SQA 工作的责任。这样的组织拥有一些方法完成这些任务，例如一个 SQA 将：

① 评审并核查开发组织符合从事软件开发标准的程度和它实施软件开发的措施和程序。

② 在项目的所有阶段，对项目活动展开独立的评审。

③ 评估半成品和最终软件产品符合企业和项目标准的程度。

④ 评估软件开发中使用的管理和工程流程（例如设计和代码的审查范围和审查内容是否足够，是否有合适的个人参与），收集和质量相关的过程数据和产品数据。

因此，参与其中的个人肩负着完成软件可靠性目标的重任，他们需要与项目的 SQA 紧密地协同工作，以避免工作的重复。要确保收集到合适的数据，协调各自的任务。

软件安全（Software Safety）：软件安全关注于特定类型的软件失效研究，安全失效也就是导致死亡或严重后果的失效。开发安全代码，评估软件安全的具体技术将在第 7.4.2 节和第 7.5.1 节中描述。

7.3 软件开发：经典的瀑布式生命周期

软件开发生命周期是一个时间阶段，从软件产品的构思开始，到产品不能再使用结束。图 7.2 给出了软件开发流程的主要阶段，它们一起构成了瀑布式生命周期。虽然各

阶段的序号、名称和为行为负责的人员有所变化，但瀑布式生命周期为理解开发流程的一般原则和它所包含的主要阶段提供了一个良好基础。

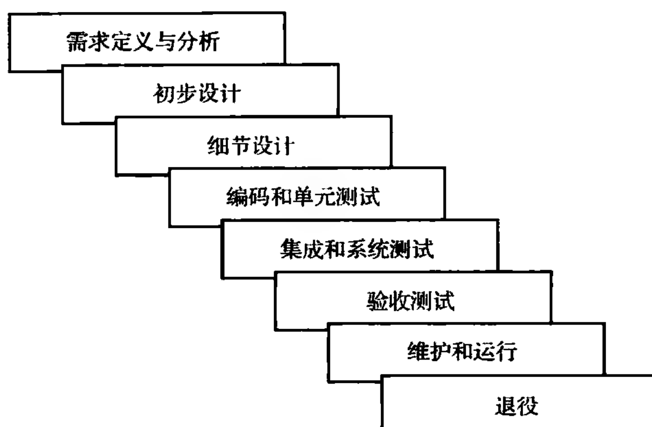


图 7.2 经典的瀑布式生命周期

通常来说，一个项目的软件生命周期不同于瀑布式生命周期模型或是它的任何数学模型。例如原则上讲，生命周期中每个阶段的最后都应该能看到一个新阶段的开始，然而在现实中明确并分析所有需求之前，初步设计通常就已开始；在定义完整的软件设计之前，编码工作就已开始。因此，实际的生命周期可能是在八个不同阶段之间的一个有序迭代。

对于小范围或通过小组（可能多于4、5个人）努力就能完成的项目，正式的软件生命周期模型（Life-Cycle Model，LCM）不是必需的。这种情况下，各小组的活动会根据项目要求从一个阶段进行到另一个阶段，组成员之间并没有相互协调。然而，一旦项目复杂性提高或组成员增加超过了一定限度，项目组织就要建立并使用软件生命周期模型，这将产生更好的开发团队，提高开发效率，并改进组工作的管理能力，生产出可靠的产品。

当定义软件生命周期模型时，需要为每个阶段确定三个任务元素：

① 将要完成的活动。

② 将要取得的产品或结果。

③ 需要面对的评审或其他里程碑。

一旦这样定义了软件的 LCM，就可以评估进度、成本以及项目相关的风险了。对于可靠性专业人员，软件的 LCM 将决定或影响：

① 可靠性工程和评估活动的时间和类型。

② 需要收集的数据和收集方法。

③ 每个阶段可能存在的可靠性评估类型。

7.3.1 各阶段描述

接下来的内容中，我们将描述软件生命周期的主要阶段，讨论每一阶段的错误或故障的来源。

1. 软件需求的定义和分析阶段

此阶段的目的是为将要开发的软件定义并用文件记载工程需求。主要成果是一个足够详细的软件需求规范（Software Requirement Specification，SRS），在项目的设计阶段，开发组织将根据此需求开展工作。需求可以是正面的，也可以是负面的。正面的需求定义所要求的（或容许的）产品运行行为，负面需求定义不可接受的产品运行行为（也可以据此定义具体的软件失效实例）。

为项目建立软件需求的方法有三种：

① 在发包（授予合同）之前，由客户指定需求，这些需求是项目的软件开发的起始点。典型的客户需求，要由客户所在组织（或许需要分包人的支持）的一个小组在用户社区的协调下制定，并把指定的需求作为后续软件开发的基础。

② 软件开发组织的初步任务是制定一个软件需求规范；在 Top-Down（自顶向下）方法中，通过在用户社区采访人群，评审政策或与所开发产品相关的功能性文件可以确定产品的用户需求、特定需求和软件的运行约束。

③ 在快速原型方法中确定软件需求的一个极小集，然后用高级语言开发可执行原型代码，以模拟所开发产品的主要功能。例如在一个管理信息系统中，开发原型代码可以描绘数据输入过程，开发视频显示终端屏幕样本可以描述用户在操作过程中将看到的东西，还需要原型化报告样本或其他输出内容。潜在用户（客户）审查原型运行的结果，然后会对性能和缺陷进行评论。将此过程迭代多次，直至判定原型产品可接受时为止。那时，原型就变成了一个实际的需求规范，也就是交付产品将要成为的模型。

对于一个软件可靠性计划来说，需求定义和分析阶段的可追踪错误是最大的损害，因为这些错误经常会传播到代码的多个区域。如果一个实时应用程序的时间需求是模糊的或是没有适当定义，最终产品的许多功能都将受到影响。在此阶段中，以下方面可能会产生错误：

① 不完整的需求：没有定义一些功能或性能需求。

② 不适宜的需求：定义的需求或许与当前硬件、计划所用硬件或操作员环境不兼容，例如一个算法定义的精度可能与输入到算法的采样率不一致。

③ 相冲突的需求：两个或更多结合在一起的需求可能不兼容，常见的例子发生在销售点系统产品中，当一个职员进入一项业务时，系统要更新交易文件的数字，并在一个特定时间为操作员返回控制权。

④ 软件需求：在与其他产品需求一致的情况下，硬件或用户的操作策略要与软件需求相兼容。

⑤ 没有适当地描述用户需求：对于除了最简单产品之外的所有产品，很难确定其用户，例如用户是产品操作者，还是操作者的上级，还是用产品来开展业务的管理者，还是产品要运行的网络的管理员？答案可能是“以上都是”。如果在需求开发时，这些人之间的冲突没有解决，那么，产品接下来的可接受性和可靠性将会是一个问题，反映在产品需求中的用户需求将不能够表达所期望的产品性能。

2. 初步设计和细节设计阶段

早期的设计工作把软件产品分解为一个层级结构，它包括计算机软件配置项（Computer Software Configuration Items，CSCI），还可以把它进一步分解为计算机软件组件（Computer Software Components，CSC）、计算机软件单元（Computer Software Units，CSU）——通常称为模块。CSCI 是软件元素的集合，为了配置管理的目的把它当作一个单元（例如 CSCI 通常在确定软件版本的那一级中）。CSU 是代码最小的可编码组，可以把它当作一个实体。在早期设计阶段，将 CSCI 的软件需求分配给若干 CSC，然后定义每个 CSC 的功能以及它所要求的输入和输出。这些信息构成了早期的软件设计。在详细设计阶段，CSC 设计规范会逐步细化，直到确定了描述产品运行元素的原始（不可分解的）CSU。

细节设计阶段应该独立于语言之外，应该允许某些非最初设计者完成程序的代码编写工作，并随着设计的进展，继续为测试活动制订计划。在早期设计阶段，测试组织建立对 CSC 集成和测试的需求；在详细设计阶段，此组织要为 CSU 集成测试和产品测试建立测试条件语句。

在此阶段，以下几个方面可能会产生错误：

① 输入数据范围错误：输入数据的允许范围可能无法反应实际应用的需求；这些错误的范围可能很大，也可能很小，因而不能用字母数字或其他字符序列来进行适当的定义。

② 不一致的数据定义：当模块在不同单元内交换信息、处理变量时，可能会出现此类问题（例如模块 A 使用的时间单位是秒，而模块 B 使用的是小时）。

③ 对算法错误的不当分析：通常，对去尾和舍入错误的分析都是正确的，然而，一个算法序列很难对输入变量范围内的错误需求进行评估。

④ 不充分的有效性检查：单个变量的有效性检查实施起来相对比较简单，但要检查输入组合的有效性，就需要更加深入地了解软件的目的（例如一个人的出生日期是不是要比死亡日期晚？一项任务的结束日期是不是在开始日期之前）。

⑤ 界面错误：一些模块会调用它们自身吗？一组模块是来自同一个菊花链吗（也就是 A 调用 B，B 调用 C，C 调用 A）？是不是有合适的输入集合传递到每个模块？

⑥ 缺乏错误恢复：当一个模块不能执行时，会发生什么情况？设计将产生灾难性后果，还是将提供一些合理的恢复行为？

3. 代码和单元测试阶段

在编码阶段，程序员把详细的软件设计转换成编程语言，此语言是在需求或设计阶段指定的。高级语言，如 C 或 Ada 中的代码被称为源代码。编译器用来把源代码转换成机器可读的目标代码。

当完成了一个单元的编码，需要对代码进行审查或校对，以确定任何设计错误、确定代码是否能实现设计、决定代码风格（例如命名约定）的一致性和标准结构。根据项目需求，这些审阅过程可能是正式的或是非正式的。尽管如此，在所有情况下的目标都是一样的：在故障变成运行中的失效之前，查明它是否存在于代码中。

执行单元测试，以核实单元的功能。在测试期间，程序员对测试条件拥有最高控制权。通常，这也是查找并及时清除故障的最后机会。在接下来的测试阶段，把单元聚集为 CSC 或 CSCI，通常，隔离并清除故障很费时间（成本也很大）。

编码是一个故障多发的过程。一个训练有素的程序单元创建者要记住代码中的所有细节，然后用耐心和毅力去实现设计，还要开发一套完整的单元测试程序。若干标准的 bug（与在书面文字中发现的语法错误一样）将会出现在这个阶段。例如：

① 丢失代码。

② 无法访问的代码。

③ 不适当或不完整的有效性检查。

④ 不完整变量和参数的初始化或重置。

⑤ 分支条件和循环的逻辑不合适。

⑥ 超出范围的计算和无限循环。

⑦ 设计文档中记载的任何失效。

4. 集成和系统测试阶段

当单元测试完成，就开始进行集成测试。此时，已单独测试的单元被组装到逻辑组合中，并进行再次测试，以演示组件满足产品需求的情况。这个过程通过结合越来越大的单元组重复进行，直到已整合 CSCI 中的所有单元为止。

集成测试完成后，执行系统测试，以演示软件产品（或许包括多个 CSCI）和硬件一起完成所有功能的情况，准备好结果产品，将其发布到生产过程。此处还要准备好产品手册、其他文档和训练材料，保证它们可用，并且和软件产品一致。

测试阶段不直接为最后产品产生代码。此阶段引入的错误和故障会影响测试项目在软件中检测失效的能力，这些错误和故障包括：

① 测试计划或程序不正确地诠释软件需求或没有追踪这些需求。

② 为测试项目（例如测试条件语句、驱动或特定数据库）所编写代码中的错误、缺陷或故障。

5. 验收测试阶段

此时，开发组停止测试，将产品转给验收测试组。验收测试组决定产品是否满足了最初的需求。指定一个验收测试计划，当其中所有测试都成功完成后，此阶段结束（不同的测试技术将在第 7.4.3 节中描述）。

6. 维护和运行阶段

如果软件需求是不变的，软件试运行或微调以改进性能时会出现一些错误，相关人员需要把焦点放在修正这些错误上。另一方面，如果软件需求根据变化的用户需求或硬件而持续改变，在此阶段要修改软件以满足运行需求，此阶段类似于一个迷你版的生命周期。

7. 退役阶段

在某个时间点，用户可能会决定不再使用软件，甚至抛弃它。由于经常性的改变，软件可能会变得不可再维护（例如文件大小不可管理或不完整、丢失）。

7.3.2 软件开发标准

DOD-STD-2167A：防御系统软件开发 1988 [DoD, 1988] 建立了“可用于整个系统生命周期的软件开发统一要求”。虽然标准没有明确指出或禁止使用任何具体的软件开发方法，但它使用了一个类似于瀑布的生命周期来标示开发阶段，并且确定了每个阶段相关的文件和审查工作。这些阶段的大部分都与 7.3.1 节中描述的相似。

MIL-STD-498：软件开发和文献编制 1994 [DoD, 1994]，取代了 DOD-STD-2167A 和其他 DoD 软件标准。虽然 DOD-STD-2167A 没有定义瀑布模型，但 MIL-STD-498 明确地给开发者提供了更多的选择。它描述了在一个或多个“构建版本”中开发软件的可能性，“要在每个构建版本中开展一些活动，而其他的活动只能在所选择构建版本中开展……直到完成所有的构建版本”。在 MIL-STD-498 中，CSCI 被分解为软件单元，这些单元可能以层级的方式相互关联，也可能不相互关联。这提供了比 DOD-STD-2167A 更灵活的方法，也能与面向对象设计更好地兼容。它提供了基于计算机的配置管理工具更灵活的使用方法。

IEE/IEC12207：信息技术标准 1996 [IEEE, 1996] 在 1998 年正式取代了 MIL-STD-498。这个标准定义了一系列过程，它覆盖了整个软件生命周期，从概念设计到退役。对于每个过程，它还定义了一个或多个信息项，这些项是过程的输入或输出内容。这个标准有三卷：12207.0-1996 描述了基本标准，12207.1-1997 是一个生命周期数据指南，12207.3-1997 是一个过程实施指南。

7.3.3 软件开发生命周期和相关成本中的错误分布

理想状况下，在生命周期开发过程的每个新阶段中，错误的数量都会减少，在单元测试中会清除大量的失效，集成测试中会清除小部分失效，甚至在系统测试中也会清除一小部分失效，在运行阶段则不清除失效。另外，在集成测试中清除的错误很可能与界面问题相关，在系统测试中清除的错误应该与软件和硬件之间的兼容性相关等。然而研究表明，一些软件项目却表现出完全不同的情况 [Neufelder, 1993]。失效数量在系统测试阶段会达到峰值，更糟糕的是它会随着阶段的推进而逐步增加。后面这种情况会在经常改变或添加需求的软件项目中出现。清除生命周期后期发现的错误可能需要很高的成本。实际上，有数据表明，当一个开发进度从生命周期模型中的一个阶段进入到另一个阶段时，修正错误的代价有一个大致为 10 的增长因子 [Neufelder, 1993]。

7.4 改进软件可靠性的技术

目前，已有一些可用于改进软件产品可靠性和安全的技术。以下将对这些技术进行简要描述。

7.4.1 可靠软件的设计

软件工程已经孕育了许多技术，这些技术帮助程序员系统地把软件设计从详细规范变成了真实的软件。本节将介绍一些软件设计的常用技术（有兴趣的读者可以参考

[Bell, Morrey and Pugh, 1992]，以获取关于这个特定主题的更多内容)。

1. 结构化编程

结构化编程直接关系到设计的清晰度。它要求设计要易于理解；要在的软件重要组成部分中构建项目结构；这些结构之间的相互关系要清晰明了，不能因为过多的细节而变得模棱两可。在结构化编程的指导原则中，首先要限制控制结构的数量。程序员要严格限制排序（一系列相继执行的声明）、判定语句（写成 if…then…else…声明）和循环（写成 while…do…声明）。图 7.3 给出了这三种控制结构，它们分别只有一个入口点和一个出口点，让软件能够在不同的抽象程度上易于更广泛人群理解。结构化编程不允许使用 GOTO，因为它们会让软件拥有多个出口点。

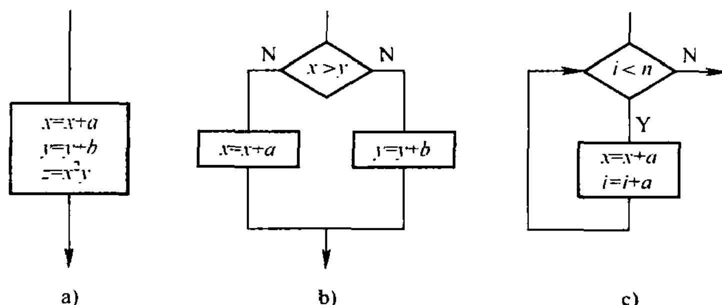


图 7.3 基本的结构化编程构架

a) 序列 b) 判断（if…then…else…） c) 循环（while…do…）

2. 设计技术

用于编程的控制结构的数量和类型有一定的局限性，它没有完整地定义程序设计，需要另外一些技术对其规范进行细化。

功能分解（Functional Decomposition）：功能分解是一种结构化编程方法，用于定义详细的软件设计和软件的总体架构。该方法自顶向下进行，就像它的名字所隐含的意思那样，它首先着重于产品需要完成的功能，其次才是数据。它优先考虑高级功能（高度抽象）。设计师决定如何使用低抽象度的功能完成这些高级功能，例如某人想起动汽车的发动机（高级功能），就要打开车门，进入车舱，插入钥匙，然后转动钥匙（四个低级功能）。一旦完全定义了某个层次的抽象功能，设计过程进入到下一个抽象层次。因此，这个技术实际上是对功能的逐步完善，设计由伪代码语言（也就是语句序列，它以动词开始，使用结构化编程认可的构造形式）表示。此过程可能会用到广度优先（每次处理一个抽象层次）和深度优先（在所有抽象层次中，只处理一个功能，然后返回，处理下一个高级功能）方法。

功能分解的主要缺点是其意义不太明确。因为固有的模糊性，它很难应用，不同的程序员使用它，必将会创建出不同的设计。

数据结构设计（Data Structure Design）：数据结构设计，也叫 Jackson 结构化编程，它面向详细程序设计的开发，它也使用伪代码语言（PseuDocode Language，PDL）。该方法和它生成的设计都由两个文件结构驱动：软件读取的输入文件结构和将

要生成的输出文件结构。根据此方法，理想的软件设计应该再现I/O文件中的数据结构。目前，此方法是最系统的设计方法。因为它明确指出了那些需要被严格遵守的步骤，它简单、理性、易于学习、有一定的规律可遵循且具有一致性：两个不同的程序员将最终写出相同的程序。然而，此方法不适用于科学计算程序，因为它的内容实质上是算法（泛函）。

3. 设计问题

程序模块化（Program Modularity）：这是软件设计中的一个关键问题。模块化是软件工程，更准确地说是软件设计中的一个关键问题。一个模块应该有多大、多复杂？程序模块之间应该有多少交互？这些问题都要在软件设计中予以考虑。从本质上讲，如果一个程序的建立是与其他程序相独立的，那么它就是一个模块。我们应该采取模块化设计，这样才能使模块设计、调试和测试更加容易，也能使维护更加便利。另外，很明显的是它允许独立地开发同一软件的不同部分，并增加重用的机会。

模块大小（Module Size）：一个极端的观点认为，模块的大小应该限制在7行代码以内，或者更小。此说法来自于心理观察，它认为人类每次只能记住七个事物。然而，如果这样做的话，将会导致软件模块之间产生巨量的交互，并会增加复杂性，进而导致改进代码可靠性的尝试失败。实际上，以往经验表明，代码的行数不能反映软件的复杂程度，因此需要制订其他的量来度量软件的复杂程度。

模块复杂性（Module Complexity）：McCabe复杂性度量（1985）把模块复杂性与决策点数量关联起来（见第7.5.2节）。McCabe断言，对任何软件模块而言，其复杂度应该不能超过11，然而此观点不足以令人相信。试想一下，理解冗长的代码的目的是多么困难的事情？即便代码只有几个决策点。

全局数据的使用：模块复杂性的特征不仅包括代码的长度和它的决策点数量，还包括全局数据。大部分全局数据（多个模块共享的数据）都对代码可靠性极为不利。使用局部数据的程序易于学习，也易于清除，也不会对剩余软件造成污染。同样的概念包括信息隐藏[㊀]和数据抽象或封装[㊁]，它们用于面向对象编程，可以实现更好的模块度（也就是说，它们提高了互换性、独立开发程度和理解性）。

内聚（Cohesion）和耦合（Coupling）：开发人员要搭建一些模块以保证模块间的交互是有限的（低耦合），大量的交互被限制在一个模块中（高内聚）。不同类型的内聚包括：偶然内聚（当模块内容任意限定）、逻辑内聚（当模块以一系列逻辑相似的功能呈现，例如打印工资和打印出生日期）、时间内聚（当模块包括需要同时完成的功能）、通信内聚（当功能根据处理相同数据分组）、功能内聚（当模块执行且仅执行一个功能）。

共享模块的设计：共享模块需要自下而上设计，以保证它们与自身内部使用的数据

㊀ 信息隐藏建立在特定对象（代码片段）的用户不应该访问对象内部的基础之上。

㊁ 数据封装是指把数据和程序（运算）包含在一个普通对象中，只有使用为对象定义的程序，才能对此对象中的数据进行修改。

相独立。

7.4.2 容错软件的设计

容错软件（Fault-Tolerant Software）源自于一个思想，即软件可接受任何数量的测试（参见第7.4.3节），允许使用任何广度的形式证法（见第7.4.4节），当然，要创建没有任何错误的软件是不可能的［Neufelder，1993；Scott，Gault and McAllister，1987］。

在意识到这个局限性的前提下，改进软件可靠性的另一种方法是：建立一个容错架构，它将使软件从失效中恢复。这种方法的昂贵代价限制了它在高风险应用程序（如控制卫星的嵌入式软件、生命关键（Life-Critical）应用程序⊖的）中的使用。

1. 程序恢复块的设计

程序恢复块（Recovery-Block）的设计如图7.4所示。软件的版本都在相同的需

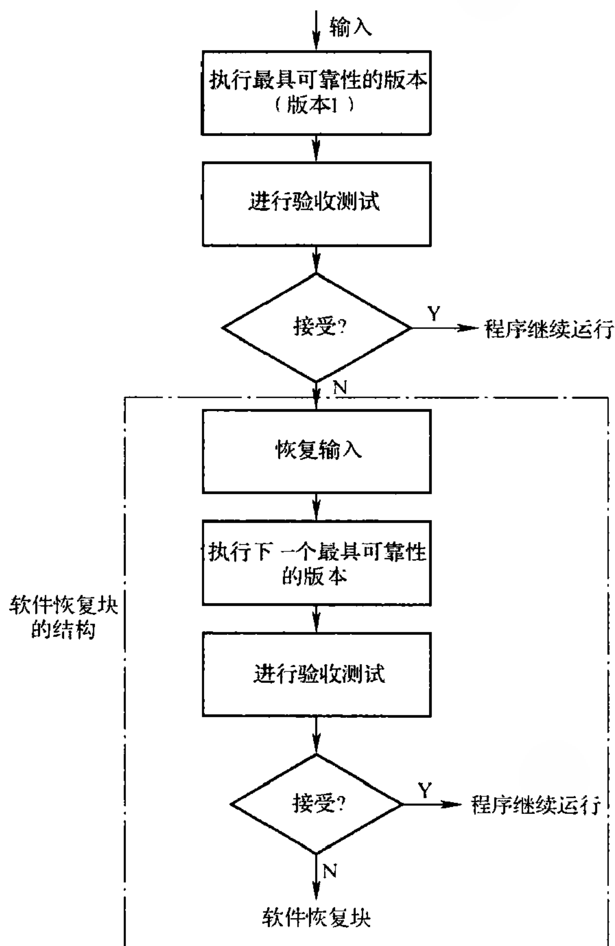


图7.4 容错设计：软件恢复块的结构

⊖ 译者注：生命关键应用程序是指一个程序的失效或故障能引起以下后果：人的死亡或重伤；设备的损失或严重损伤；环境危害。

求下开发。假定这些程序段是独立开发的，那么它们同时（在相同的输入下）失效的可能性几乎可以忽略不计，当然也要假设软件需求是正确的。把软件的版本按照可靠性从高到低排序（每个程序段的测试强度能够为排序提供基础），最具可靠性的版本将优先执行。一旦输出计算完成，就开始进行验收测试。如果输出没有通过验收，就进入一个程序恢复块。恢复过程包括三个步骤：恢复输入，用恢复的输入执行下一个最可靠的软件，提交输出到相同的验收测试，如果接受，进入下一个程序恢复块，依次类推。此设计的两个主要弱点是输入的恢复和验收测试，因此需要小心处理它们。如果程序恢复块的设计比单独的程序段更可靠，就非常有必要对验收测试软件进行彻底的测试。

2. *N* 版本编程

在 *N* 版本编程中，要独立开发软件的几个版本。*N* 个程序是平行执行的（见图7.5）。在各版本执行完成以后，要对输出进行比较。如果至少有两个程序的输出相同，那么就声明输出正确，接受输出，过程继续。这样的设计没有程序恢复块的那两个弱点。然而，它不适用于那些会产生多个正确输出的应用程序（如计算地图两点之间路径的程序），并且还会歧视受到舍入错误影响的正确答案。

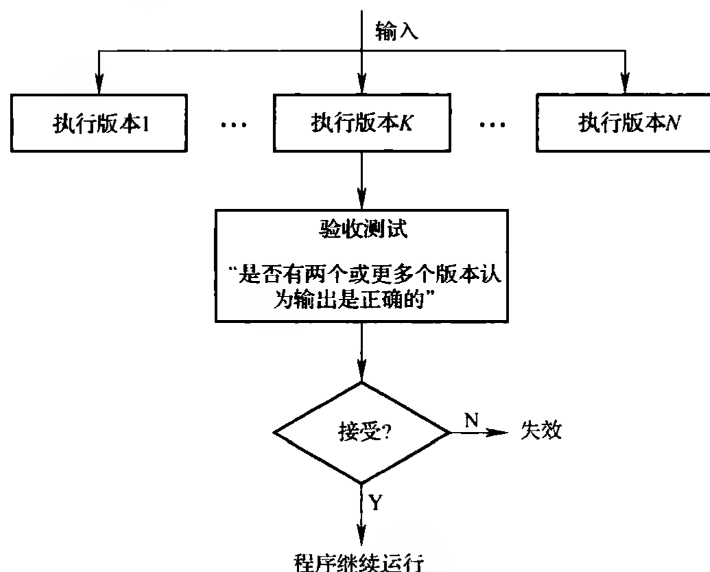


图7.5　容错设计：*N* 版本编程

3. 一致性恢复块

一致性恢复块（Consensus Recovery Block）（见图7.6）结合了程序恢复块和 *N* 版本编程的特征，它试图消除以上两种设计的弱点。一致性恢复块要求开发一个程序的 *N* 个不同版本以及一个验收测试的投票程序。根据可靠性程度，将软件的不同版本进行分级。执行所有的版本，并把输出提交到投票程序。如果没有一致同意，陆续把每个输出提交到验收测试，以测试其可靠性。只要有一个输出通过测试，马上停止过程，软件继续进行计算。一致性恢复块比前面提到的两个容错设计都更可靠。

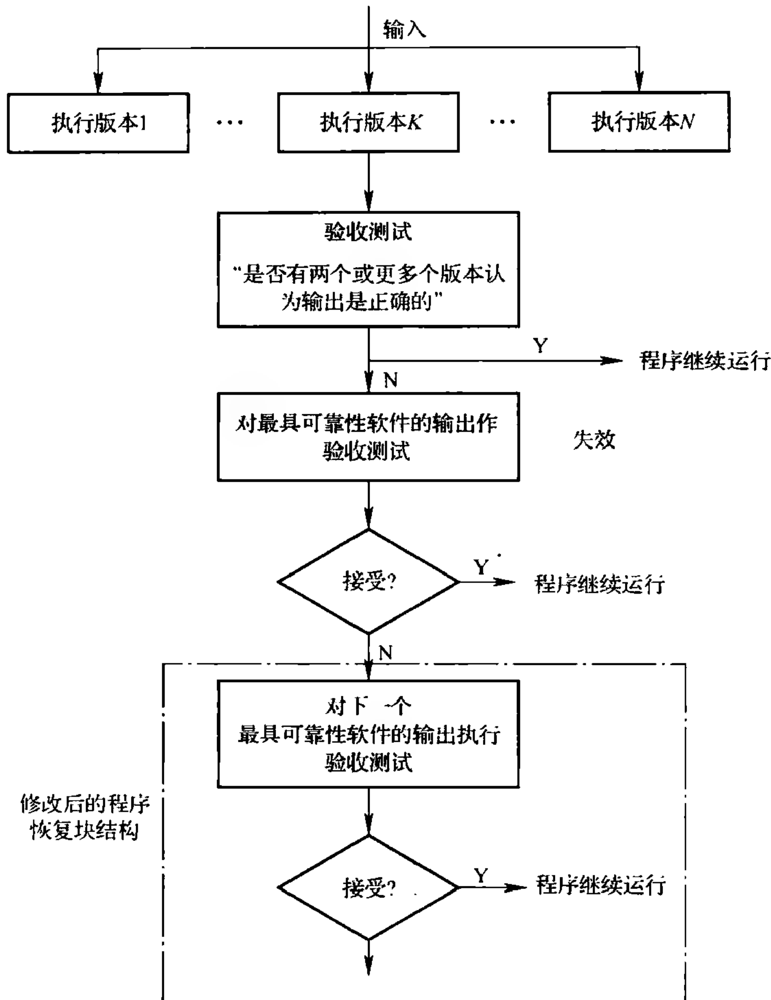


图 7.6　容错设计：一致性恢复块

7.4.3　测试

软件测试是一个程序执行过程，其目的是找出错误。然而，无论多少测试条件语句也不能保证软件完全没有错误。即使验证一个小程序，也可能需要大量的测试条件语句。

一些原则和策略可以帮助程序员更有效地对程序进行测试。对于测试的简要介绍，感兴趣的读者可以参考文献 [Myers，1979]，也可以参考文献 [Beizer，1984]，其中有关于测试技术更全面的报告。此节将介绍一些测试方法，主要是模块测试和集成测试。

1. 黑盒测试和白盒测试

测试策略的两个大类是黑盒（Black-Box）测试和白盒（White-Box）测试。白盒测试要用到程序结构知识，让测试条件语句尽可能多地涵盖编程中的逻辑路径。黑盒测试把程序看作一个黑匣子，完全不考虑编程结构，通常它关注程序的输入和输出（见第

7.4.3.2节)。表7.2中列出了一些特殊的白盒、黑盒测试技术，我们将在下一节对这些技术进行详细讨论。

表7.2 模块测试：白盒测试技术与黑盒测试技术

黑 盒	白 盒
等价划分 边界值分析	语句覆盖 判定覆盖 条件覆盖 判定/条件覆盖 多条件覆盖

整个程序测试的良好覆盖范围，要求把白盒、黑盒测试结合起来使用。实际上，因为前者需要更深的程序知识，通常为程序开发员所用，它们可能会忽略掉程序中的瑕疵。此外，白盒测试不测试程序与需求的一致性。

在实践中，要优先用黑盒测试方法开发测试条件语句，用白盒测试方法开发补充测试条件语句。

2. 模块测试：白盒测试策略与黑盒测试策略

逻辑覆盖测试（Logical-Coverage Testing）：表7.3定义了不同的白盒逻辑覆盖测试技术，它们用来测试程序M（见图7.7）。我们将此作为一个例子，使用条件覆盖测试技术推导出与测试条件语句对应的设定。这样做的目标是至少让每个条件在每个判定语句处执行一次。

表7.3 逻辑覆盖测试技术

逻辑覆盖	描 述	程序M的测试情况
语句覆盖	程序中每个语句至少执行一次	A=2，B=0
判定覆盖	执行每个判定语句的任意一个输出	A=2，B=0 A=0，B=1，X=0，C=2
条件覆盖	每个判定语句的每个条件至少执行一次	A=2，B=0，C=3，X=2 A=0，B=1，C=2，X=0
条件/判定覆盖	每个条件、每个判定语句的任意一个输出的结果至少执行一次	A=2，B=0，C=3，X=2 A=0，B=1，C=2，X=0
多条件覆盖	每个条件组合至少执行一次	A=2，B=0，C=3，X=2 A=2，B=0，C=2，X=2 A=2，B=1，C=2，X=0 A=2，B=1，C=3，X=0 A=0，B=0，C=2，X=2 A=0，B=0，C=3，X=2 A=0，B=1，C=2，X=0 A=0，B=1，C=3，X=0

我们的目的是每个判定语句的每个条件至少执行一次。所测试程序拥有：

① 两个判定语句——也就是“A > 1 or C = 2 and B = 0”和“A = 2 or X > 1 or C = 3”。

② 六个条件：A > 1，C = 2，B = 0，A = 2，X > 1，C = 3。

```
M:PROCEDURE(A.B.C.X):
IF(((A >1)|(C=2)&(B=0))THEN DO.
            X=X/A:
            END:
IF((A=2)|(X> 1)|(C=3))THEN DO:
            X=X+1:
            C=X+A:
END
```

图 7.7　程序 M 的代码片段

选择 A > 1 or A≤1，C = 2 or C≠2，B = 0 or B≠0，A = 2 or A≠2，X > 1 or X≤1，C = 3 or C≠3 产生的结果，例如第一种情况可能包括 A = 2（满足 A > 1 and A = 2），B = 0（满足 B = 0），C = 3（满足 C≠2，C = 3），X = 2（满足 X > 1），第二种情况可能包括 A = 0（满足 A≤1 and A≠2），B = 1（满足 B≠0），C = 2（满足 C = 2，C≠3），X = 0（满足 X≤1）。

等价划分（Equivalence Partitioning）：此技术的目标是找出尽可能多地涵盖不同输入测试条件语句的最小集合。为了实现此目标，它把程序的输入域划分为一个有限数量的等价类——也就是一些子域，在这些子域中，无论输入条件是什么，程序都会有相同的输出。因此，使用等价划分来设计测试案例包括两个步骤：

① 确定等价类（那些有效的输入类属于有效等价类，无效的输入就是无效等价类）。

② 定义测试案例（在同一时间段内，尽可能多地覆盖所有有效等价类，每次用一个有效等价类来设计测试案例）。

假设把用来输入数组大小的程序记为 A，软件需求所定义 A 的范围为 10 ~ 200。此时，只有一个有效等价类“10 ≤ A ≤ 200”，两个无效等价类是“A < 10”和“A > 200”。因此，将产生三种测试情况，例如 A = 50，A = 9，A = 201。等效划分的假设条件是：两个输入值属于同样的等价类，这些类能引发程序同样的反应——换句话就是，这些输入值与程序行为等价。

图 7.7 中的软件程序拥有三个声明：“X = X/A”、“X = X + 1”和“C = X + A”，两个判定语句：“A > 1 or C = 2 and B = 0”和“A = 2 or X > 1 or C = 3”。每个判定语句都可能有两个输出：“true”和“false”。判定语句“A > 1 or C = 2 and B = 0”有三个条件：“A > 1”、“C = 2”和“B = 0”。

边界值分析（Boundary Value Analysis）：边界值分析是一项用于生成条件语句的技术，它的基础是测试条件比其他条件更能引发软件故障。这些测试条件要探索输入和输入等价类的边界。如果以 A 的大小为例，使用边界值分析的测试情况将包括一个输入集合，其大小是 A = 9，A = 10，A = 200 和 A = 201。

3. 集成测试

如在第 7.3.1 节中所简要介绍的那样，只要模块测试一完成，马上就启动集成测试（Integration Testing），逐步把模块结合到测试界面中。把不同元素结合到最终软件产品需要一定的策略，它将影响模块实际的编码和测试顺序、测试成本、生成测试条件语句的成本。非累进测试（Nonincremental Testing）首先会单独地检查所有不同的模块——

且完成这些测试行为，就把软件完全组装起来，并进行界面错误测试。累进测试（Incremental Testing）首先测试一个模块，再把它和第二个还没有测试的模块结合起来；然后对组合模块进行测试；然后把下一个模块添加到组合模块中，测试过程继续进行。累进测试有两种执行方式：自顶向下和由下而上。由下而上测试要使用驱动模块[⊖]，而自顶向下测试要求广泛使用桩模块（Stub）（见图7.8）。表7.4列出了以上两种累进测试方式的优缺点。

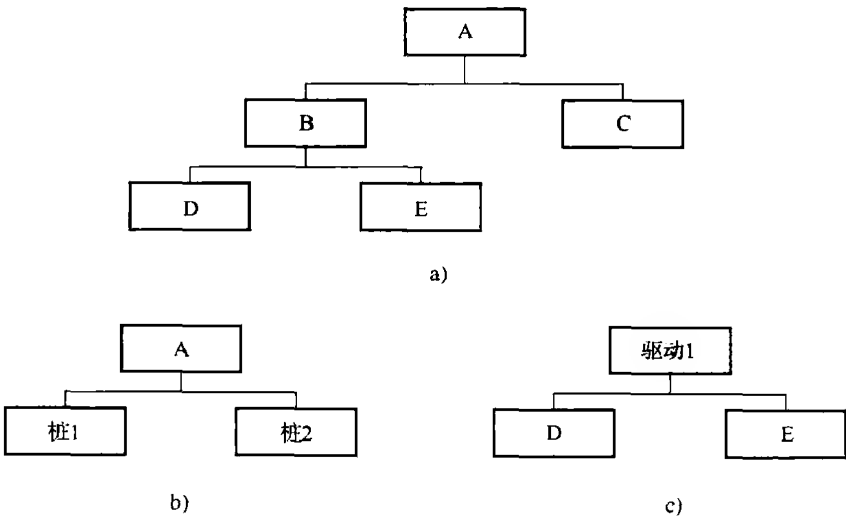


图7.8　累进式的集成测试

a）处于集成测试阶段的程序　b）自上而下集成测试的第一步　c）由下而上集成测试的第一步

表7.4　自顶向下和由下而上测试的比较

	优　点	缺　点
自顶向下测试	在高级软件中能快速找到错误位置 只要添加了I/O功能，就能更容易地表达测试情况 可以在早期论证骨骼程序	必须编写桩模块 桩模块可能会比较复杂 在添加I/O功能之前，在桩模块中表达测试条件更难 不能创建测试条件，或比较难创建测试输出的观察更难 设计和测试可能会重叠 对某个模块测试的完成可能会推迟
由下而上测试	在低级软件中能快速找到错误位置 测试条件比较容易创建 容易观察测试结果	必须有驱动模块 在最后一个模块添加之前，程序不能以完整状态呈现

⊖ 驱动模块（Driver Module）是需要进行编码的模块，它驱动测试情况语句或把试验情况传送到测试中。必须要编码一个桩模块，模拟被测试模块在执行期间调用程序的行为。

7.4.4 形式化方法

1. 形式化规范方法 [Neuhold and Paul, 1991]

形式化规范方法（Formal Specification Method）包括一个详细的项目需求规范，它使用形式语言——不是传统编程中使用的自然语言，而是一个有精确定义语法和语义的“正式”替代语言。数学就是该语言的一个例子。

下面这个例子描述了把逻辑作为一个规范语言的使用过程。其内容是一个提款机的运行过程，用户需要从他们的银行账号里面取款或者存款、查看账户状态等。在取款过程中，提款机首先要筛选访问，只承认那些带有磁卡且有正确的个人身份证账号（Pin）的用户服务需求。相同磁卡重复三次尝试失败后将导致退出磁卡。用户的可借记次数取决于账号的收支平衡状态、银行允许的信用额度以及每次或一周的取款额度限制。这些约束条件可在一个形式语言中予以定义。

x: card number;

y: pin;

a: access variable ($a(x)=1$, access is allowed to x); and

$\forall x, \forall y$ such that (pin(x) $= y$ and attempts (x) 3 implies $a(x)=1$).

这些约束条件的所有组合形成了一个形式规范——提款机软件行为的形式规范。

1973 年，IBM 维也纳研究实验室提出了维也纳开发方法（Vienna Development Method，VDM）。这是一个广泛流传的形式规范方法，它使用工业应用已采用的逻辑。VDM 的校验建立在一阶逻辑的基础之上。

形式规范强迫在定义需求时清晰、明确，但也为误解和歧义性保留了一定的空间。通常，这些都是使用自然语言所要关注的问题。形式规范可以为自身所用，也可以与证明工具结合起来论证程序满足需求的情况（见第 7.4.4 节）或作为自动代码生成的基础（此处不考虑）。

2. 形式验证

形式验证（Formal Verification）[Galton，1992] 对形式规范进行推理，以论证程序满足需求的程度。

下面这个一阶逻辑的例子：由少量的额外特征增广为一个证明系统。本节将使用此证明系统：Geo 是一个软件模块，用来计算一个几何级数的 n 个开头项 $s=1+q+q^2+q^3+\cdots+q^k+\cdots+q^{n-1}$，程序的输入为 q 和 n。假定 q 和 n 都是正整数，输出为 s（因此 $s>0$），后面会给出此程序（见图 7.9）。此证明系统建立的基础是文献 [Hoare，1969] 中所提出的传统逻辑的扩展形式。在启动一个程序之后，此证明工具可用来确定假设。假设是关系到程序输入变量的先决条件，它能引导至后置条件的正确集合，后置条件与输出相关。把 $\{P\}$ 记作先决条件集合，$\{Q\}$ 记作后置条件集合，S 是软件的记号。

```
Procedure geo(q,n:integer; var s,k:integer)
s:=1
k:=1
While k<n do
begin
s:=s+q^k, k:=k+1     ] L(code internal to the while·· do· loop)
end
end.
```

图 7.9　Geo 程序的代码

我们想要表达的就是 $\{P\}\ S\ \{Q\}$。

传统逻辑的扩展包括：

① 一个公理，它用来描述赋值语句的结果：$\{P[x'/x]\}\ x:=x'\ \{P\}$，其中，$P[x'/x]$是逻辑的性质 P，其中所有出现的 x 都已被 x'代替。此规则表明，如果把 x'分配给 x，则变量 x'的所拥有性质都为变量 x 所拥有。

② 一些用来指示如何组织两个先决条件和后置条件的规则：

规则1：如果 $\{P\}\ S\ \{Q\}$，且 $\{Q\}$ 包含 $\{R\}$，那么 $\{P\}\ S\ \{R\}$；

规则2：如果 $\{P\}\ S\ \{Q\}$，且 $\{R\}$ 包含 $\{P\}$，那么 $\{R\}\ S\ \{Q\}$。

③ 一条用来指导如何组织两个软件性质的规则：

规则3：如果 $\{P\}\ S\ \{Q\}$，且 $\{Q\}\ S\ \{R\}$，那么 $\{P\}\ S;\ S'\ \{R\}$。其中，$S;\ S'$指 S'在 S 之后运行。

④ 一条用以描述如何处理“while… do…”循环的规则：

规则4：如果 $\{P$ 或 $B\}\ S\ \{P\}$，那么，当 B 执行 S $\{$非 B 和 $P\}$ 可得 $\{P\}$。

我们将证明 $\{q>0\}$ Geo $\{s>0$ 且 $q>0\}$。证明过程从程序最后的语句开始，逆向贯穿不同的程序语句。

证明：使用赋值公理，$\{s>0$ 且 $q>0\}$ $k:=k+1$ $\{s>0$ 且 $q>0\}$。再次使用赋值公理，$\{s+q^k>0$ 且 $q>0\}$ $s:=s+q^k$ $\{s>0$ 且 $q>0\}$。在包含 $\{s+q^k>0$ 且 $q>0\}$ 的 $\{s>0$ 且 $q>0\}$ 上使用规则2，则产生 $\{s>0$ 且 $q>0\}$ L $\{s>0$ 且 $q>0\}$。也就是说，$\{s>0$ 且 $q>0\}$ 是 while…do…循环的一个变量。

然后使用规则4，$P=\{s>0$ 且 $q>0\}$ 和 $B=\{k<n\}$ 会产生 $\{s>0$ 且 $q>0\}$，当 $k<n$ 时，执行 L $\{k>=n$ 且 $s>0$ 且 $q>0\}$，由规则1可知，它即是 P。

再次使用赋值公理，$\{1>0$ 且 $q>0\}$ $s:=1;\ k:=1$ $\{s>0$ 且 $q>0\}$。然而，$\{1>0$ 且 $q>0\}$ 等价于 $\{q>0\}$，因此 $\{q>0\}$ Geo $\{s>0$ 且 $q>0\}$。

这个简短的例子说明了如何基于一阶逻辑使用形式证明系统，以论证软件产品可以在所有的输入条件下满足需求。在这方面，形式方法与传统的测试技术有着巨大的差异，后者仅能用来应付有限数量的输入条件，这些条件比输入域有更多的采样、更少的灵活性。当然，主要问题是确定软件需要验证的性质并对这些性质进行证明。证明过程可以自动执行，此类工具正在缓慢地开发。即便是全自动执行，该方法仍未被全部利用起来，因为只有很少程序员知道如何使用它们。其形式方法很难理解，因为它们本身很复杂，需要经过大量的练习才能掌握。

7.4.5 软件开发过程成熟度

曾经有一个关于应用新的软件方法和技术将获得一定的生产效率和质量的承诺，但一直没有实现。直到20年之后，工业和政府组织才意识到他们的根本问题所在：根本无法管理软件过程 [DoD, 1987]。在一个缺少宽广组织的软件过程之下，重复的成功结果通常依赖于在下一个项目使用同样的个体。这种方法不能为企业组织提供长期的软件质量和可靠性改进基础。

1986年，软件工程研究所（Software Engineering Institute, SEI）开始为软件开发组

织开发一个能力成熟度模型（Capability Maturity Model，CMM）［Paulk et al，1993］。CMM 为所有组织描述了软件开发过程成熟度的 5 个级别。高成熟度级别意味着更好的可预测性、低风险、软件质量和可靠性的增加。CMM 为那些想要改进其开发、维护软件流程的组织提供了一个指南，也为采购组织提供了指南，用以评估与某个特定组织签订软件项目合同的风险。

成熟度的 5 个级别是：

① 初始的：软件开发过程的特征是临时的、只定义了少数过程、项目产出很难预测。

② 可重复的：建立了基本项目管理过程，以追踪开发成本、日程和功能。管理过程可能会因项目不同而有所改变，但管理控制是标准化的，它可判定当前状态在整个项目寿命中的地位。对于相似的项目，开发组织极有可能重复以前实现的性能级别。

③ 已定义的：软件的管理和工程活动过程都是有记载的、标准化的，它集成到一个拥有宽广组织的软件过程。所有项目的软件开发和维护都使用一个已记载且已一致通过的组织流程版本。

④ 有管理的：收集了关键过程的详细措施。过程和产品都可被量化理解和控制，使用详细的软件度量。

⑤ 最优化的：从过程和测试启发的思想、技术中得到的定量反馈，使得对过程进行持续改进成为可能。

CMM 在每一个成熟度级别确定了“关键过程区域”（Key Process Area，KPA），它包括组织和项目的目标和一些活动，这些活动在一个区域中推进、记载、验证过程成熟度。第 2 级的 KPA 包括需求管理、项目计划、项目追踪和监管、软件分包管理、软件质量保障和软件配置管理。

SW-CMM 评估在 2000～2004 年执行，在 2005 年 1 月提交报告给 SEI，报告中有超过 1900 个组织。这些评估表明 10% 的组织处于第 1 级，42% 的组织处于第 2 级，31% 的组织处于第 3 级，8% 的组织处于第 4 级，9% 的组织处于第 5 级。

SW-CMM 得到了扩展和更新，已能够处理系统工程问题［Bate et al，1995］。目前，SW-CMM 已被 CMMI（Capability Maturity Model Integration，能力成熟度模型集成）取代。CMMI 也被称作系统工程能力成熟度模型（SE-CMM）。

7.5 软件可靠性评估技术

在此节中，我们列出了若干众所周知的技术，它们用来对软件可靠性进行定性和定量评估。

7.5.1 软件分析方法

失效模式及影响分析（Failure Mode and Effect Analysis）和故障树分析（Fault-Tree Analysis）（参见本书的第 4、5、9 章）也可应用于软件开发工作。

1. 失效模式及影响分析（FMEA）

FMEA 是一个由下而上的方法，它连续假定产品的每个单元失效，并在所有可能出

现的有害结果中追踪这些失效。FMECA（失效模式、影响及危害度分析，Failure Mode and Effect Criticality Analysis）包括每个失效模式的危害度级别，此级别建立在失效概率和（或）危害严重度之上。这些技术在硬件中有着众多的应用案例，然而应用到软件的案例却仍然非常稀少，可能是因为很难确定软件组成部分的失效模式。在系统级别（那些软件和硬件结合的产品），这些技术已经得到了成功应用。

2. 故障树分析

故障树从失效影响（也称作故障树顶事件或危害）开始逆向进行，以确定某事件（单个组件的失效）的根源。

图7.10为程序 *M*（见第7.4.3节）的故障树分析图。在代码层（最健康的层），

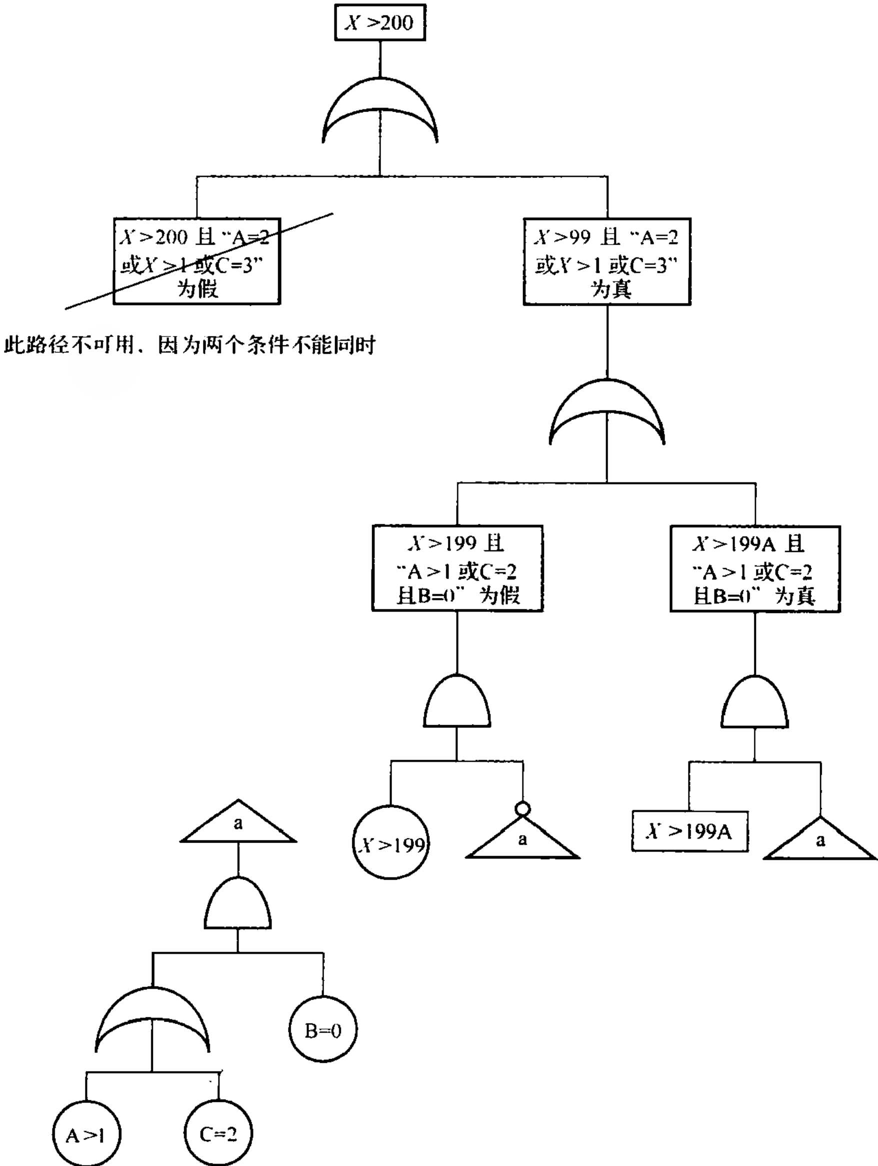


图7.10　程序 *M* 的软件故障树分析

仅对某一顶事件“$X>200$”进行分析。换句话说，我们想在输出变量 $X>200$ 时，确定输入条件的情况。如果 M 是某个控制发动机转速软件的一部分，$X>200$ 可能会超出发动机的设计极限，从而破坏发动机。在软件的不同路径中，从一个满足 $X>200$ 的值开始逆向使用，此方法确定不同变量的值以及导致这种不可接受输出的参数。有时候，在代码层进行分析是不可承受的，特别是对超过了 100000 行的代码进行分析。此方法仅限于应用到模块或子模块层。然后再向下到代码层。代码层是保证安全的关键区域 [Leveson and Harvey，1983]。

7.5.2 软件度量

软件度量是一个软件产品、开发过程拥有某给定属性的量化指标。度量经常用于评估软件开发工作状态和趋势，以评估从生命周期的一个阶段进行到另一个阶段的风险。

1. 需求量度：完备量度的规范

通常，软件需求阶段不产生代码，但有可能会有原型活动或尝试重用其他项目的一些代码的行为发生。在此阶段，可用主要的软件量度来表示规范的大小、完整性和稳定性以及软件生命周期进入下一阶段的风险。此时，可用的量度是：

① M_1 = 已充分定义或规定的需求。

② M_2 = 没有定义或规定的需求。

③ M_3 = 需求总量 = M_1+M_2。

④ M_4 = 定义测试需求的需求。

M_3 对时间的曲线表示需求的确定是否稳定，M_1/M_3 对时间的曲线表示定义已知需求过程的完整性。最后，M_4/M_3 对时间的曲线表示测试要求的确定情况。

如果一开始 M_3 就不稳定或 M_1/M_3 和 M_4/M_3 没有达到某个合理的阈值（即 75%），那么，规范中的不确定因素就产生技术风险。在进入到下一阶段前，要对此技术风险进行管理审查。

2. 设计阶段的量度

除了用以确定错误源的错误数据分析之外，其他量度在此阶段也有一定的作用。

① 完整性量度（Completeness Measure）：用来确定与规范和设计相关的技术风险区域。

② 复杂性量度（Complexity Measure）：用来评估把设计转换成代码的潜在难度。

③ 缺陷密度量度（Defect Density Measure）：用来评估设计阶段的可靠性成长情况。

完整性指标（Completeness Indicator）：完整性指标 [Department of the Air Force，1987] 提供了洞察软件规范适当程度的方法。它的使用可以从软件开发过程的需求阶段开始，且在整个软件开发生命周期中有效。由完整性指标的部分内容确定的值可用来确定技术风险区域，也可用来评估产品等级需求转换到规范和后续设计的程度。

此指标的输入内容是软件规范需求和设计成熟度。这些内容可直接从需求分析、设计和代码审查中获得。鉴于此指标的用途，要将功能和需求做等价考虑。最初，可以从软件规范说明评审（Software Specification Review，SSR）或初步的设计审查（Preliminary Design Review，PDR）中获取数据，把它作为每个输入参数的指标。随着软件进入到设

计、编码和单元测试阶段，要对这些数据更新。对于有多个 CSCI 的产品来说，需要对每个 CSCI 进行独立计算。

完整性指标的输入内容包括：

P_1 = 没有充分定义或具体指定（SSR）的功能数量；它与需求分析阶段的 M_2 相同；

P_2 = 功能总数（SSR），与需求分析阶段的 M_3 相同；

P_3 = 没有定义的数据项（PDR）；

P_4 = 数据项的总数量（PDR）；

P_5 = 已定义，但未使用的功能数量（PDR）；

P_6 = 已定义功能总数量（SSR）（$P_6 = P_2 - P_1$）；它与需求分析阶段的 M_1 相同；

P_7 = 被已定义功能引用，但未定义（PDR）的功能；

P_8 = 被已定义功能引用的功能总数量（PDR）；

P_9 = 未使用任何条件或选项的决策点数量（PDR）；

P_{10} = 决策点总数量（PDR）；

P_{11} = 未处理的条件选项数量（PDR）；

P_{12} = 条件选项总数量（PDR）；

P_{13} = 所调用参数与已定义单数不一致的例程调用数量（PDR）；

P_{14} = 例程调用总数量（PDR）；

P_{15} = 未设定的条件选项数量（PDR）；

P_{16} = 已设定，但没有用选项处理的条件选项数量；

P_{17} = 需要设定的条件选项的数量（PDR）（$P_{17} = P_{12} = P_{15}$）；

P_{18} = 没有目标的数据引用次数（PDR）；

一旦收集了这些输入，计算下面这些指数：

已定义功能的满意度（C_1）：$C_1 = (P_2 - P_1)/P_2$；

已定义的数据库引用或项（C_2）：$C_2 = (P_4 - P_3)/P_4$；

所使用的已定义功能（C_3）：$C_3 = (P_6 - P_5)/P_6$；

已定义的引用功能（C_4）：$C_4 = (P_8 - P_7)/P_8$；

在决策点使用的所有条件选项（C_5）：$C_5 = (P_{10} - P_9)/P_{10}$；

在决策点处使用过，且正在处理的所有条件选项（C_6）：$C_6 = (P_{12} - P_{11})/P_{12}$；

所有调用例程参数，这些参数与已调用例程定义所定义参数一致（C_7）：$C_7 = (P_{14} - P_{13})/P_{14}$；

所有设定的条件选项（C_8）：$C_8 = (P_{12} - P_{15})/P_{12}$；

所有跟随设定条件选项的进程（C_9）：$C_9 = (P_{17} - P_{16})/P_7$；

已有目标的所有数据项（C_{10}）：$C_{10} = (P_4 - P_{18})/P_4$。

然后，完整性就可以计算为这 10 项的加权总和

$$\text{COMPLETENESS} = \sum_{i=1}^{10} w_i C_i \tag{7.1}$$

其中，w_i 是每个部分的权重（其值在 0 到 1 之间），所有权重总和为 1。因为每个 C_i

也处于0和1之间，所以完整性量度通常也在0到1之间，值越大，表明规范越完整。

虽然在项目早期可能很难决定 P_1 到 P_{18} 这些单独的输入值，但估计其组成部分的值却应该比较容易，因为它们都是分式形式。例如在项目早期，虽然 P_1 和 P_2 的精确值不知道，但有工程经验的人可以估计 C_1 为0.1或0.2。

开发一个成功产品，需要确定每个部分在公式中的重要性，也就是权重 w_i，它用来计算可靠性。它是规范复杂性、应用类型、重用代码量、其他计划或应用因素的函数。例如在需求分析和早期设计阶段，把 w_i 设定为1；在其他阶段，把它设为0是比较合适的。

虽然完整性是一个复杂的量度，但可以在设计阶段的早期对其进行计算，并可以在整个开发过程中（例如在初步和细节设计的议定时）周期性地更新。在这些指标的组成部分中观测到的值以及值的走向，可以用来确定规范和技术风险区域的稳定性。

复杂性量度：文献［McCabe，1985］第一次介绍了循环复杂度（Cyclomatic Complexity）概念（IEEE），它用来决定模块操作图中某个模块的结构复杂度。模块操作图可以从设计阶段的信息中构建，一旦编码开始，则从程序代码中构建。

一个模块的强连接图包括四个原始的元素（“强连接”指每一个节点都可以从其他节点到达。在出口节点和入口节点之间增加一条边界，就可以实现强连接）。如果：

① N 为节点数（程序语句的有序组）。

② E 为边界数（节点之间的程序流程）。

③ SN 是断开节点（有超过一个的边界它发出）的总数。有 N 个退出路径的节点为 SN 贡献一个 $N-1$。例如当一个模块包括一个有 N 个路径的判定，如一个CASE语句有 N 种情况，此语句为 SN 贡献一个 $N-1$ 值。

④ RG 是区域的数量（由边界围成的区域，且边界没有交叉）。

那么，模块操作图的循环复杂度（C）计算为

$$
\begin{aligned}
C &= E - N + 1 \\
&= SN + 1 \\
&= RG
\end{aligned}
$$

例如图7.11所示的一个模块操作图有8个节点（$A \sim H$）、12条边、5个区域，其循环复杂度为

$$
\begin{aligned}
C &= \text{边数} - \text{节点数} + 1 = 12 - 8 + 1 = 5 \\
&= \text{断开阶段数} + 1 = 4 + 1 = 5 \\
&= \text{由路径围成的区域数} = 5
\end{aligned}
$$

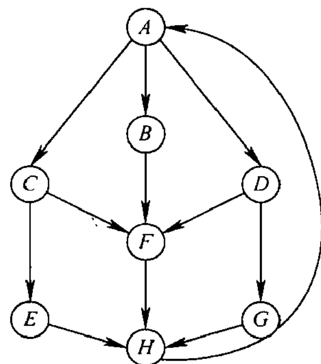


图7.11　有8个节点（$A \sim H$）、12条边、5个区域的模块操作图

这些区域是 *ACEHA*、*ACFHA*、*ABFHA*、*ADFHA* 和 *ADGHA*。

从物理上讲，循环复杂度表示穿过模块的唯一路径数。模块的复杂度可表示为对其进行正确编码、测试、修正和改进所要付出的努力。因此，在设计阶段对模块复杂度进行估计，有利于确定模块区域，这样就可以对其进行进一步的分解或简单的设计修正。

虽然很难快速确定一个硬性的规定，但通常当复杂度超过10或11就表示需要重新

设计，以便使其简化。

缺陷密度（Defect Density）［IEEE，1989a］：这个量度要求建立缺陷严重度类型，并收集以下数据：

D_i——独特缺陷的总数，独特缺陷是在某个特定严重性等级，在第 i 个设计审查中发现的缺陷；

I——审查的总数；

KSLOD——在设计阶段，设计说明中的源代码行数，以千计。

那么，设计的累积缺陷率（Cumulative Defect Ratio For Design，DD）计算为

$$DD = \sum_{i=1}^{I} D_i / KSLOD \tag{7.2}$$

在此度量方法中有一些二义性，因为一个较小的值可能意味着：要么是一个好的产品，要么是一个不好的审查过程。如果缺陷密度计算值比可比项目中曾有过的值高，那么就要对开发过程进行重新审查，以判断原因是人员训练的不足还是开发措施的不足或判断决策需求是不完整或模糊的。

在此情况下，比较合适的解决方法可能是延迟开发，直到采取了恢复行为。如果缺陷密度的计算值比可比项目中曾有过的值低，就需要对审查过程和方法重新审查；如果问题已确定，对审查程序进行额外的训练或修正，这样或许能解决问题；如果审查过程充分，就有理由推断：开发过程的这些阶段产生出了低缺陷的产品。

3. 代码和单元测试阶段的量度：缺陷密度［IEEE，1989a］

缺陷密度参数的第二种形式适用于此阶段。再一次建立缺陷严重度分类，并追踪收集的数据：

D_i——独特缺陷的总数；

I——审查的总次数；

KSLOD——已审查源代码的总行数，以千计。

那么，代码的累积缺陷率（CD）计算为

$$CD = \sum_{i=1}^{I} D_i / KSLOD \tag{7.3}$$

7.5.3　软件可靠性模型

1. 软件可靠性模型的分类法

过去数年中，人们已开发了很多软件可靠性模型。关于细节描述，参见文献［Musa，Iannino and Okumoto，1987］。本节全面概述了现有的可靠性模型，给出并评论了这些模型的假设基础，描述了最常用软件可靠性模型（Software Reliability Model，SRM）中四种模型的细节，并介绍了它们固有的局限性。

过去，软件可靠性模型有一些不同的分类方法。Musa、Iannino 和 Okumoto 认为，最好的 SRM 模型建立在故障排除过程中的 Markovian 公式之上[⊖]。在测试开始时，软件

⊖ Markovian 模型为过程的独立历史状态之间的转换描述了随机过程（参见第2章）。

中出现的故障数量与每个故障失效概率密度函数的分布[⊖]不同。有限失效类型的模型建立在一个假设之上，即在时间 0 或至少在平均时间处，软件中故障的精确数量是可知的。无限失效类型的模型也建立在一个假设之上，即失效的数量是无限的。有限失效类型分为两种类型：二项分布型（时间 0 处，故障的精确数量可知）和 Poisson 型（故障的平均数量可知）。根据模型中使用的每故障失效强度分布，可将这两种模型分类（见图 7.12）。

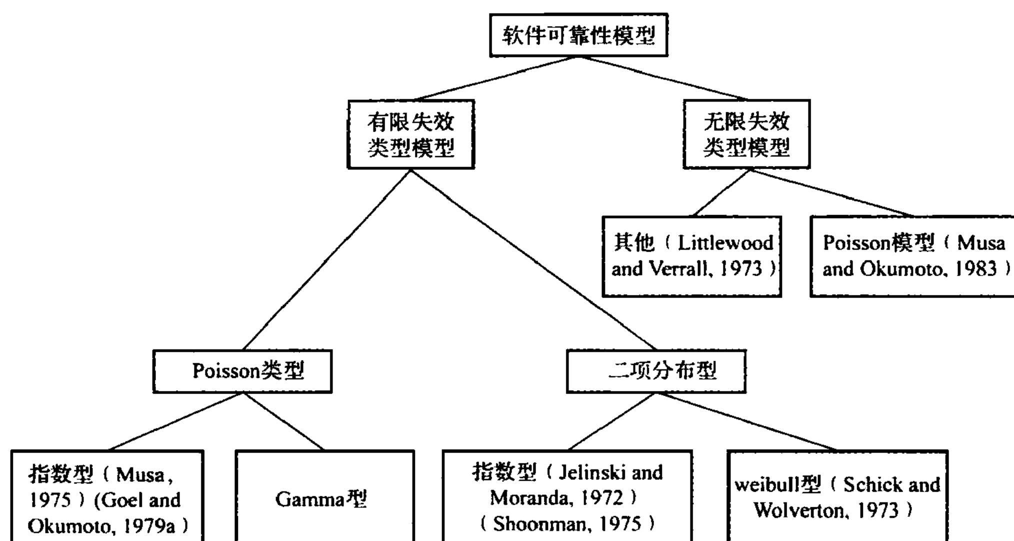


图 7.12　软件可靠性模型分类法

此处给出的分类法是文献［Goel，1985］所提出分类法的一个普及版本。它包括大部分现有的 SRM，也为选择软件可靠性模型以满足特定应用给出了指导。大部分现有的 SRM 都可以归为以下四类中的一类：

① 失效间隔时间类型（Time-Between-Failure Category）：它包括计算两次失效间的模型。

② 失效次数类型（Failure Count Category）：它关注于特定时间间隔内，故障或失效的次数。

③ 故障播种类型（Fault Seeding Category）：它包括一个模型，用以计算程序中在时间 0 时，通过外部故障播种产生的故障数量。

④ 基于输入域的类型（Input Domain-Based Category）：它包括一个模型，用以在一个众所周知的程序输入的运行分布中随机取样，并作为测试案例时，评估程序可靠性。"净室方法（Clean Room Methodology）"是此方法在工业环境（软件工程实验室，NASA）中的试探性实施方法［Basili and Green，1993］。可靠性估计从执行期间观测到的失效数量中获取。

⊖ 每故障概率密度函数 $f_a(t)$：$f_a(t)\mathrm{d}t$，是故障"a"在时间 t 和 $t+\mathrm{d}t$ 之间引发软件失效的概率。

表7.5列出了每种模型类型的关键假设基础以及每个类型的代表性模型。每个模型的具体的附加假设在表7.6中列出。表7.7检查了某些假设的有效性（它们是某类型的一般假设，还是某特定模型的附加假设）。

表7.5 每种模型类型的关键假设基础

关键假设	特定模型
失效间隔时间模型： ① 失效间隔时间相互独立 ② 与每次故障出现的概率相等 ③ 故障相互独立 ④ 在修复中没有新的故障出现	① Jelinski-Moranda 的 De-Eutrophication 模型（1972） ② Schick-Wolverton 模型（1973） ③ Goel-Okumoto 不完整调试模型（1979） ④ Littlewood-Verral Bayesian 模型（1973）
失效次数模型： ① 测试间隔相互独立 ② 间隔期的测试呈非齐次分布 ③ 在非重叠时间间隔内检测到的故障数量相互独立	① Shooman 指数模型（1975） ② Goel-Okumoto 非齐次 Poisson 过程模型（1979）Goel 通用非齐次 Poisson 过程模型（1983） ③ Musa 执行时间模型（1975） ④ Musa-Okumoto 对数 Poisson 执行时间模型（1983）
故障播种模型： ① 播种的故障在程序中随机分布 ② 原有和播种的故障被检测到的概率相等	Mill 播种模型（1972）
基于输入域的模型： ① 输入资料分布未知 ② 使用随机测试（随机选择输入） ③ 输入域可划分为等价类	① Nelson 模型（1978） ② Ramamoorthy-Bastani 模型（1982）

表7.6 与每个软件可靠性模型相关的具体假设

模型种类代表	具体假设
失效间隔时间模型： Jelinski-Moranda De-Eutrophication 模型 Schick-Wolverton 模型 Goel-Okumoto 不完整调试模型 Littlewood-Verral Bayesian 模型	在时间0处有 N 个故障；立即清除检测到的故障；两次失效的时间间隔中的故障率[①]与仍然存在的故障数量成比例 同上，故障率是仍然存在的故障数量和最后一次失效前的时间的比例 同上，但即便检测到故障，也不一定清除 失效的初始数量未知，失效间的时间服从指数分布，故障率服从 gamma 分布
失效次数模型： Shooman 指数模型 Goel-Okumoto 非齐次 Poisson 过程模型 Goel 一般非齐次 Poisson 过程模型 Musa 执行时间模型 Musa-Okumoto 对数 Poisson 执行时间模型	假设与 JM 模型中相同 失效累积数量服从非齐次 Poisson 过程（NHPP）；失效率对时间呈指数降低 与 Goel - Okumoto NHPP 的假设相同，但失效率尝试优于替代测试试验的结果，后者表明显示失效率先随时间上升，然后降低

（续）

模型种类代表	具体假设
故障播种模型： Mill 播种模型	与 JM 模型中的假设相同 与 Goel - Okumoto NHPP 模型中的假设相同，但此模型中考虑的时间是执行时间
基于输入域的模型： Nelson 模型 Ramamoorthy-Bastani 模型	测试案例的结果为其他输入提供一些关于程序行为的随机信息，这些输入接近于测试中使用的输入

① 软件故障率为 $z(t)=f(t)/R(t)$，其中，$R(t)$ 为时间 t 和 $f(t)=\mathrm{d}R(t)/\mathrm{d}t$ 处的可靠性函数。软件失效率（强度）为 $h(t)=\mathrm{d}\mu(t)/\mathrm{d}t$，其中，$\mu(t)$ 为时间 t 内发生的失效累积的平均值。

表 7.7　一些软件可靠性模型假设的有效性

假　设	假设固有的局限性
失效间隔时间相互独立	只有测试情况的变化真正随机（永远不可能）时才成立
检测到的故障会立即被清除掉	故障会立即被排除；通常会批量地修复故障；然而，只要后续测试避开了故障所在的路径，此假设就无效
在故障移除过程中不会有新的故障出现	一般来讲，这种情况是不可能的
失效率随测试时间降低	很多情况都近似于此
所有故障的失效率与仍然存在的故障数量成正比	如果选择的测试情况能保证对代码不同部分进行相等概率的测试，此假设具有一定的合理性
把时间用作度量失效率的基础	通常，时间是度量失效率的良好基础；如果不是这样，模型不适用于其他单元
在给定失效间隔下，失效间的失效率增加	通常不是这样的，除非测试强度增加
测试是运行使用的代表	通常不是这样的，因为通常选择错误易发情况进行测试（见第 7.4.3 节）

软件开发过程具有环境依赖性。因此，即便在某功能或产品的测试中能看到合理的结果，但仍然可能在后续的测试中出现另外的情况。用户仍然要做出关于假设的适当程度以及某个模型的适用性的最终决策。

为某个特定应用程序选择一个 SRM，以下方法较为实用：首先确定软件可靠性模型的所属类型；然后在给定种类中评估特定模型适用于应用程序的程度。实际上，如果模型属于前两种类型——也就是失效间隔时间类型或失效次数类型中的一个，只需要做一个选择就可以了。如果情况是这样的，根据软件的开发过程和环境选择一个可靠性模型。收集一些失效数据，如失效次数、性质、发生失效的时间、严重度、隔离故障，以确定修复所需的时间。把失效累积次数和失效强度描述为时间的函数，从所收集数据中获取这些模型的不同参数，并使用模型预测未来行为。如果未来行为与模型预测相符，

保留此模型。

2. Jelinski-Moranda 模型

文献［Jelinksi and Moranda，1972］建立了这个最早的可靠性模型，它假设：

① 一个程序中的所有故障都有同等概率在测试中引发失效。

② 故障率和仍然存在的故障成比例。

③ 在测试或调试期间，软件中没有出现新的缺陷。

最初，模型假定在每次失效后，只有一个故障被清除，但此模型的一个扩展——Sukert 模型允许清除多个故障。

在此模型中，软件的故障率为两次失效间隔常数，它也是第（$i-1$）和第 i 次失效时间间隔内的故障率

$$Z_i(T)=\phi[N-n_{i-1}](i=1,2,\cdots,m) \tag{7.4}$$

其中，N 是软件中故障的初始数量，ϕ 是一个比例常数，n_{i-1}是从第一个（$i-1$）间隔内清除的故障累积数量。

N 和 ϕ 的极大似然估计可由以下公式计算：

$$\sum_{i=1}^{m}\frac{1}{N-n_{i-1}}-\sum_{i=1}^{m}\phi x_{1i}=0 \tag{7.5}$$

和

$$\frac{n}{\phi}-\sum_{i=1}^{m}[N-n_{i-1}]x_{1i}=0 \tag{7.6}$$

其中，x_{1i}是第（$i-1$）和第 i 次失效的间隔时间长度，n 是目前为止清除的错误数。只要计算了 N 和 b，剩余的错误数就可以计算为

$$N(remaining)=N-n_i \tag{7.7}$$

到下一次软件失效的平均时间为

$$\mathrm{MTTF}=\frac{1}{(N-n_i)\phi} \tag{7.8}$$

可靠度为

$$R_{i+1}(t|t_i)=\exp[-(N-n_i)\phi(t-t_i)](t\geqslant 0) \tag{7.9}$$

其中，假定软件在 t_i 时刻发生第 i 次失效，$R_{i+1}(t|t_i)$是软件在间隔［t_i，t_{i+1}］中 t 时刻的可靠度。

3. Musa 基本执行时间模型（BETM）

文献［Musa，1975］第一次提出了此模型。它假定失效以非齐次 Poisson 过程发生。失效强度单位是每一个中央处理单元（CPU）的失效时间。这样，就把失效事件和软件使用的处理器时间连接了起来。在 BETM 中，无论修正了第一次还是第 N 次失效，失效强度函数化简后仍然为常数。

失效强度是失效经历的函数：

$$\lambda(\mu)=\lambda_0(1-\mu/\nu_0) \tag{7.10}$$

其中，$\lambda(\mu)$ 是失效强度（CPU 每小时在 μ 失效时的失效），λ_0 是初始失效强度（在 $\tau_e=0$ 处），μ 是在执行时间的平均失效经历次数，τ_e、ν_0 是在有限时间内期望的失

效总数。

然后，从当前失效强度 λ_P 到目标强度 λ_F 需要发生的失效次数计算为

$$\Delta\mu = \frac{\nu_0}{\lambda_0}(\lambda_P - \lambda_F) \tag{7.11}$$

要达到此目标的执行时间为

$$\Delta\tau = \frac{\nu_0}{\lambda_0}\ln(\lambda_P/\lambda_F) \tag{7.12}$$

实际中，有三种方法可以计算 ν_0 和 λ_0：

① 使用相似软件的经历，模型就可以在测试之前使用。

② 描绘实际测试数据，建立或更新前面的计算值。描绘失效强度执行时间：数据拟合直线的 y 截距是 λ_0 的计算值；描绘失效强度失效次数：数据拟合直线的 x 截距是 ν_0 的计算值。

③ 使用测试数据建立一个极大似然估计。文献［Musa et al，1987］详细描述了这个方法。

Musa 还提出了一个方法，用以把执行时间转换为日历时间。日历时间的组成需要一个基础，即可用资源制约了一个日历日（Calendar Day）中实际可用的执行时间总量。

4. Musa-Okumoto 对数 Poisson 执行时间模型（LPETM）

文献［Musa and Okumoto，1983］第一次提出了对数 Poisson 执行时间模型（Logarithmic Poisson Execution Time Model，LPETM）。在 LPETM 中，失效强度计算为

$$\lambda(\mu) = \lambda_0\exp(-\theta\mu) \tag{7.13}$$

其中，θ 是失效强度衰变参数，λ、μ 和 λ_0 与 BETM 使用的相同。参数 θ 表示每次失效经历的失效强度变化。此模型假定：第一次失效的修复对失效强度减小有很大影响，以后每次修复所受影响呈指数减小。

在 LPETM 中，不需要 v_0 值。从当前失效强度 λ_P 到目标强度 λ_F，预期一定出现的失效次数为

$$\Delta\mu = (1/\theta)\ln(\lambda_P/\lambda_F) \tag{7.14}$$

达到此目标的执行时间为

$$\Delta\tau = \frac{1}{\mu}\left[\frac{1}{\lambda_F} - \frac{1}{\lambda_P}\right] \tag{7.15}$$

根据前面的经验，在此等式中绘制测试数据，以图形化估计或为数据制订一个最小二乘法拟合可以计算 λ_0 和 θ 的值。

5. Mills 故障播种模型［IEEE，1989b］

通过播种过程可以计算仍然存在于程序中的缺陷数量，播种过程假定一个有代表性缺陷类型的齐次分布。此度量中的变量为 N_S，它是已播种的故障数量；n_S 是已发现的播种故障数量；n_F 是无意播种被发现的故障数量。

在播种前，要对一个故障进行分析，以决定将在代码中出现的故障类型以及它们的相对发生率。把一个独立检测器插入到代码 N_S 故障中，这些故障是预期本地故障的代

表。在审查（或测试）中，要确定已播种和未播种的故障。发现的播种故障和本地故障数量可用来为所考虑故障类型计算剩余故障数量。除非已发现一些播种故障，否则无法计算这个量。本地（未播种的）故障的极大似然估计为

$$NF = n_F N_S / n_S \tag{7.16}$$

例如某类型的20个故障已播种。那么，就会发现40个属于此类型的故障：16已播种，24未播种。$NF=30$，剩余故障的值为 $NF(remaining) = NF - n_f = 6$。

6. Nelson 输入域模型

在 N 次运行，K 次失效的经历中，Nelson 模型（1978）获取一个可靠度估计值 R。如果输入的 N 集合从运行剖面中随机抽取（即这些输入的分布概率，它取代实际的输入内容，进入到程序中），那么 $(1-K/N)$ 就是 R 的一个无偏估计值。

7. 衍生软件可靠性模型

根据前面介绍的基本模型，我们可以为拥有多个程序段的应用程序建立模型，例如用以评估容错软件设计可靠性的模型，在软件集成阶段评估一组已组装模块可靠性的模型。

把一个 N 版本容错设计软件作为例子。假定 N 个不同版本完全独立，如果满足以下三个条件中的一个，产品将经历一次失效设计：

① 所有输出不一致；把错误标为 E_1。

② 产生相同的错误输出（E_2）。

③ 投票程序不能实现其功能（E_3）。

如果忽略错误 E_1 和 E_2，设计就等效于一个多硬件冗余。如果 $p(V_n=c)$ 是版本 n 正确执行的概率，$p(V_n=I)$ 是版本 n 失效的概率，$n=3$，则

$$\begin{aligned} p(E_1) = {} & p(V_1=I)p(V_2=I)p(V_3=I) + p(V_1=C)p(V_2=I)p(V_3=I) + \\ & p(V_1=I)p(V_2=C)p(V_3=I) + p(V_1=I)p(V_2=I)p(V_3=C) \end{aligned}$$

其中，$p(V_3=C) \approx R_{vi}$（如果把可靠度计算为成功运行的次数，它就是版本 i 的可靠度）。

如果投票程序失效（即抛弃两个或更多同时出现的正确输出），$p(VP=I)$、$p(E_1)$ 变为

$$\begin{aligned} p(E_1) = {} & p(V_1=I)p(V_2=I)p(V_3=I) + p(V_1=C)p(V_2=I)p(V_3=I) + \\ & p(V_1=I)p(V_2=C)p(V_3=I) + p(V_1=I)p(V_2=I)p(V_3=C) + \\ & \{p(V_1=C)p(V_2=C)p(V_3=C) + p(V_1=C)p(V_2=C)p(V_3=I) + \\ & p(V_1=I)p(V_2=C)p(V_3=C) + p(V_1=C)p(V_2=I)p(V_3=C)\} \\ & p(VP=I) \end{aligned}$$

对于已知成本，可以用此模型评估出最具可靠性的设计。

文献［Littlewood，1979］明确地把软件结构（也就是模块）纳入考虑范围之内，它利用一个半 Markovian 过程，为模块间控制的改变建模。给定模块的失效率可以从应用到模块的基本可靠性模型中获取，它可以用于界面失效建模。此模型可以用来研究集成过程。

8. 对现有软件可靠性模型的评论

软件可靠性模型已经受到广泛关注。第一类反对者主要是对概念的反对。软件具有纯粹的决定性，而硬件行为则是部分随机的。一旦定义了程序的输入集合，程序将正确运行或者失效，没有“程序将失效的概率”之类的事情存在。当然，这种争论并没有什么依据。尽管如此，就像在其他模型中一样，概率在软件可靠性模型中是用来表现“不确定性”的。程序和将要运行的输入都具有不确定性，它们都不是故障的发生位置，也不具有故障的性质。

第二类反对者的观点建立在一个事实之上，即软件可靠性模型来自于硬件可靠性模型。仅对后者稍微做了些修改，就用在一些软件特性上（例如软件不会磨损，也就不会遵循浴缸曲线）。大部分此类模型的主要缺陷在于：它们建立在有限数量，且有时是有问题的假设之上，而不是建立在更深的理论知识如软件的“热水力学方程（Thermal-Hydraulics Equation）”[1]之上。

正如在第 7.5.3.1 节中讨论的那样，在过去若干年中，人们已开发、测试、验证、废弃了一定数量的模型，大部分模型都建立在软件失效数据之上。通常，它们仅限于在后期阶段的测试[2]（集成、产品和验收测试阶段）中使用。如果一个软件项目完全从零开始（没有相比较软件的可用历史），对于项目早期阶段的管理来说，这些模型将毫无用处。

7.6 总结

软件不仅仅是代码，它还包括程序、过程、规则与计算机系统运行相关的文件编制。固件是一种特殊的软件形式，它包括计算机程序和驻留在各种内存中的数据。在运行期间，这些都不会被计算修改。

软件可靠性反映了在特定条件下、特定时间段内，一个程序完成所需求功能的能力。软件失效是开发过程中发生错误的结果，反过来它将导致代码中出现故障。

软件质量关注于软件产品的众多特征，包括可靠性。因此，负责软件质量的组织、个人在达成软件可靠性的过程中将扮演重要角色。软件安全专注于生命关键应用程序，这些软件承受不起失效。要达成软件安全，需要开发极其可靠的软件。

软件开发组织可利用软件生命周期模型定义一个项目的软件开发过程。这需要确定相关活动、产品、审查或其他开发过程中需要完成的时间表，还要收集相关数据。为了给软件提供长期产品质量和可靠性改进的基础，这些组织需要使用类似于 SEI 的 CMM 程序，以管理并改进它们的工作流程和未来软件开发工作的可重复性。

在软件生命周期中，我们可以使用若干技术来减少或消除最终产品中的潜在失效。

---

[1] 一些模型，如 Littlewood 结构模型 [Littlewood，1979]，已经能更好地描述、理解软件的性质了。

[2] 注意：如果可靠性增长已通过形式验证达成，那么将不再需要我们介绍的模型。就我们所知，关于形式化方法对软件可靠性的影响，还没有开发出能考虑此问题的可靠性模型。

这些技术包括：容错设计、不同阶段的测试、形式化方法。每一项技术都有自己的优点、局限性和影响其应用范围和深度的成本。容错设计的成本比较高，这些成本必须由应用情况决定。同样，软件可靠性仍将建立在精确的表决系统之上。无论测试组织所展现出的测试技术如何，测试越详尽越好，目前还没有全自动化的测试技术。形式化方法较复杂，且要求人员经过高级训练，才能有效地应用它们。

在开发过程中，我们可以用一些量度来评估产品的可靠性特征和进行到下一个开发阶段的风险。评估可能要建立在从生命周期中收集的失效数据之上，它可用来计算可靠性和失效强度，还能计算达到特定失效强度级别所需要的时间。尽管如此，这些评估还是局限于在软件开发生命周期的集成、系统和验收测试阶段使用。除非软件开发组织已经从以往开发工作累积了一套统一的数据，那么对于项目早期阶段的管理，这些评估模型没有任何帮助。

参考文献

Basili, V., and S. Green. 1993. The evolution of software processes based upon measurement in the SEL: The cleanroom example. University of Maryland and NASA1GSFC, draft.

Bate, R. et al. 1995. A systems engineering capability maturity model, version 1.1 (CMU1SEI-95-MM-033). Pittsburgh, PA: Software Engineering Institute.

Beizer, B. 1984. Software system testing and quality assurance. New York: Van Nostrand Reinhold.

Bell, D., I. Morrey, and J. Pugh. 1992. Software engineering: A programming approach. Upper Saddle River, NJ: Prentice Hall.

DoD (Department of Defense). 1985. Defense system software development, DOD-STD-2167, Washington, D. C.

———. 1987. Report of the defense science board task force on military software, Office of the Under Secretary of Defense for Acquisition. Washington, D. C.

———. 1994. Software development and documentation. MIL-STD- 498, Washington, D. C. Department of the Air Force. 1987. Software quality indicators. Air Force Systems Command, AFSC Pamphlet 800-14.

Galton, A. 1992. Logic as a formal method. Computer Journal 35 (5).

Goel, A. L. 1983. A guidebook for software reliability assessment. Rep. RADC TR-83-176.

———. Software reliability models: Assumptions, limitations, and applicability. IEEE Transactions on Software Engineering SE-11 (12): 1411.

Goel, A. L., and K. Okumoto. 1979a. A time dependent error detection rate model for software reliability and other performance measures. IEEE Transactions on Reliability R28: 206.

———. 1979b. A Markovian model for reliability and other performance measures of soft-

ware systems. Proceedings of the National Computer Conference, New York 48.

Hoare, C. A. R. 1969. An axiomatic basis for computer programming. Communications of the ACM 12: 576.

IEEE (Institute of Electrical and Electronics Engineers). 1983. IEEE standard glossary of software engineering terminology. ANSI1IEEE Std. 729.

———. 1989a. IEEE standard dictionary of measures to produce reliable software. IEEE Std. 982. 1-1988.

———. 1989b. IEEE guide for the use of IEEE Standard Dictionary of Measures to Produce Reliable Software. ANSI/IEEE Std. 982. 2-1988.

———. Standard for information technology. IEEE/IEC12207, 1996.

Jelinski, Z., and P. Moranda. 1972. Software reliability research. In Statistical computer performance evaluation, ed. W. Freiberger. New York: Academic Press.

Leveson, N. G., and P. R. Harvey. 1983. Analyzing software safety. IEEE Transactions on Software Engineering SE-9: 5.

Littlewood, B. 1979. Software reliability model for modular program structure. IEEE Transactions on Reliability R-28: 3.

Littlewood, B., and J. L. Verrall. 1973. A Bayesian reliability growth model for computer software. Applied Statistics 22: 332.

McCabe, T. 1985. Structural testing. Columbia, MD: McCabe and Associates.

Mills, H. D. 1972. On the statistical validation of computer programs. Rep. 72-6015. Gaithersburg, MD: IBM Federal Systems Division.

Musa, J. D. 1975. A theory of software reliability and its application. IEEE Transactions on Software Engineering SE-1: 312.

Musa, J. D., A. Iannino, and K. Okumoto. 1987. Software reliability. New York: McGraw-Hill.

Musa, J. D., and K. Okumoto. 1983. A logarithmic Poisson execution time model for software reliability measurement. Proceedings 7th International Conference Software Engineering, Orlando, FL.

Myers, G. J. 1979. The art of software testing. IBM Systems Research Institute. New York: John Wiley & Sons.

Nelson, E. 1978. Estimating software reliability from test data. Microelectronic Reliability 17: 67.

Neufelder, A. M. 1993. Ensuring software reliability. New York: Marcel Dekker.

Neuhold, E. J., and M. Paul. 1991. Formal description of programming concepts. IFIP International Federation for Information Processing, Laxenburg, Austria.

Paulk, M. C. et al. 1993. Capability maturity model, version 1.1 (SEI-93-TR-024). Software Engineering Institute, Pittsburgh, PA.

Ramamoorthy, C. V., and F. B. Bastani. 1982. Software reliability: Status and perspectives. IEEE Transactions on Software Engineering SE-8: 359.

Schick, G. J., and R. W. Wolverton. 1973. Assessment of software reliability. Paper presented at llth Annual Meeting German Operational Research Society, DGOR, Hamburg, Germany; also in Proceedings of Operational Research Physica-Verlag, Wirzberg-Wien.

Scott, R. K., J. W. Gault, and D. G. McAllister. 1987. Fault tolerant software reliability model-ing, IEEE Transactions on Software Engineering SE-13: 5.

Shooman, M. L. 1975. Software reliability measurement and models. Proceedings of the Annual Reliability and Maintainability Symposium, Washington, D. C.

第 8 章 失效模式、机理及影响分析

本章介绍一种名为失效模式、机理及影响分析（Failure Modes，Mechanisms，and Effects Analysis，FMMEA）的方法，它用来确定潜在失效模式、机理以及它们产生的影响。FMMEA 通过确定高优先度的失效机理来帮助人们制定相应的行动计划，以减小这些机理所产生的影响，通过这种方式，它提升了失效模式与影响分析（Failure Modes and Effects Analysis，FMEA）和失效模式、影响及危害性分析（Failure Modes，Effects，and Criticality Analysis，FMECA）的价值。FMMEA 所用的关于失效原因及其对应的失效机理的知识，有助于提升产品开发的有效性和经济性。本章介绍了如何应用 FMMEA 进行电子电路板组件的分析。

8.1 引言

竞争激烈的市场迫使制造商寻找一些更为经济的手段，以改进产品开发过程。尤其是工业企业，他们希望通过一个有效的方法了解那些随着时间的推移可能会影响产品性能的潜在失效。一些企业正在使用或被要求使用一种称为失效模式及影响分析（FMEA）的技术来实现此目标，但其中的大部分企业并不完全认同这种做法。

20 世纪 50 年代，Grumman 飞机制造公司首先把 FMEA 作为一种正式的方法提出，并用它来分析海军战机飞行控制系统的安全性。从 20 世纪 70 年代到 90 年代，人们编写了各种军事、专业社会标准和流程，用来定义并改进 FMEA 方法［Bowles，2003；Kara-Zaitri，Keller，Fleming，1992；Guidelines for Failure Mode and Effects Analysis for Automotive，Aerospace，and General Manufacturing Industries，2003］。1971 年，电子工业协会（EIA）可靠性委员会发布了“失效模式及影响分析（Failure Mode and Effects Analysis）”。1974 年，美国国防部颁布了 Mil-Std 1629，即“执行失效模式、影响及危急度分析的程序（Procedures for Performing a Failure Mode，Effects and Criticality Analysis）”，经过数次修订，它已成为基本的系统分析方法。1985 年，国际电工委员会（IEC）提出了 IEC812“系统可靠性分析技术——失效模式及影响分析程度（Analysis Techniques for System Reliability—Procedure for Failure Modes and Effects Analysis）”。20 世纪 80 年代后期，汽车工业开始应用 FMEA 方法。1993 年，由来自 Chrysler、Ford 和 GM 组成的供应商质量要求工作组通过 QS 9000 过程把 FMEA 引入质量手册中。1994 年，汽车工程师学会（SAE）发布了 SAE J-1739“设计中的潜在失效模式及影响分析和制造与装配过程中的潜在失效模式与影响分析（Potential Failure Modes and Effects Analysis in Design and Potential Failure Modes and Effects Analysis in Manufacturing and Assembly Processes）”，它为准备实施 FMEA 提供了一个通用指南。1999 年，作为国际汽车特别工作组一员的 Daim-

ler Chrysler、Ford 和 GM 一致通过新的国际标准“ISO/TS 16949”，此标准包括 FMEA，并最终在 2006 年取代了 QS 9000。

FMEA 已作为一个六西格玛工具在众多行业中得到了应用。它适用于各种应用情况，如系统 FMEA、设计 FMEA、过程 FMEA、设备 FMEA、功能 FMEA、界面 FMEA 和细节 FMEA。虽然使用目的和术语会因行业类别的不同而有所差异，但所有的 FMEA 过程的主要目标都是在开发过程早期预测问题，并防止或最小化这些问题所产生的影响 [SAE Standard，2002]。

失效模式、影响及危害性分析（FMECA）是 FMEA 的一个扩展形式，它是一种用于评估潜在失效模式的发生率（Occurrence）及危急度（Criticality）的技术。现在，FMEA 和 FMECA 这两个术语可以互换使用 [Bowles and Bonnell，1998；Bowles，2003]。FMEA 还是一种六西格玛工具 [Franceschini and Galetto，2001]，六西格玛机构引用了它的某些形式。FMEA 方法的基础是一个分层方法，它用来确定潜在失效模式如何影响产品。FMEA 分析需要一个跨职能团队的参与，此团队拥有分析产品的整个生命周期的能力。图 8.1 是一个典型的设计 FMEA 工作表。

系　统＿＿＿＿　　　潜在失效模式及影响分析　　　FMEA编号＿＿＿＿
子系统＿＿＿＿　　　(FMEA 设计)　　　准备人＿＿＿＿
部　件＿＿＿＿　　　　　　FMEA日期＿＿＿＿
设计团队＿＿＿＿　　　关键日期＿＿＿＿　　　更正日期＿＿＿＿
核心小组＿＿＿＿　　　　　　第　/　页

											行动结果				
产品/功能	潜在失效模式	潜在失效影响	Sev	潜在失效原因	Prob	当前设计约束	Det	RPN	建议	责任和目标完成日期	所采取的行动	新 Sev	新 Occ	新 Det	新 RPN

图 8.1　FMEA 工作表

失效机理是导致失效的某些特定应力共同作用的过程，这些应力包括物理、电、化学和机械应力 [Hu et al，1993]。在分析和报告过程中，FMEA 或 FMECA 都不确定失效机理和模型。为了了解并避免失效，需要确定与失效机理相关的那些主要应力（机械、热、电、化学、辐射），正是这些主要应力暴露了失效。了解失效机理产生的原因和影响，有助于产品的设计和开发过程，这些过程包括虚拟鉴定、加速试验、根源分析和寿命消耗监测。

在虚拟鉴定（Virtual Qualification）中，失效模型用来分析估计产品的失效时间分布。没有产品相关主要失效机理和运行条件的信息，产品的虚拟鉴定就没有意义。对于加速试验设计，相关人员需要了解可能与运行条件相关的失效机理。只有具备了失效机理的知识，相关人员才能设计出合适的试验（应力级别、物理结构和持续时间），从而能够激发相关机理引起的失效，且不会导致伪失效。

所有的根源分析技术，包括因果图和故障树分析都要求参与人员了解可能影响失效事件发生的条件。失效机理分析也影响着失效根源假设和其验证过程。了解失效机理，并了解影响这些机理的应力的相关知识，对于产品的寿命消耗监测来说是非常重要的。产品的实际空间和收集、传递数据的可用接口约束了实际中可用于产品的传感器数量。

为了确保收集合适的数据，并在产品状态监控中使用这些数据进行产品剩余寿命的估计，必须对产品失效机理进行优先排序。

传统的FMEA和FMECA没有阐述用以分析产品失效的失效机理的关键问题。为了弥补此缺陷，人们提出了失效模式、机理及影响分析（FMMEA）方法。FMMEA过程结合了FMEA模板的系统性和“可靠性设计”的理念及知识。除了收集并应用于FMEA过程的信息，FMMEA还把应用条件、特定应用的持续时间、主动应力和潜在失效机理的知识结合了起来。在产品的设计和验证中，需要结合产品的预期应用条件，使用合适的评估模型对潜在的失效机理进行逐一分析。接下来的章节将详细介绍FMMEA方法。

8.2 失效模式、机理及影响分析方法

FMMEA是一个系统化的方法，它用来确定潜在失效模式的失效机理和模型，并对它们进行优先度排序。高优先度失效机理决定了运行应力、需要在设计中考虑或控制的环境和运行参数等。

FMMEA的基础是了解产品需求和产品物理特征（以及它们在生产过程中的变化）之间的关系，了解产品材料和载荷（在应用条件下的应力）之间的相互影响，了解材料和载荷对使用条件相关的产品失效敏感度的影响。这就要求寻找用来量化评估失效敏感度的失效机理和可靠性模型。图8.2是实施FMMEA的步骤。在下面的小节中，我们将对这些步骤进行详细介绍。

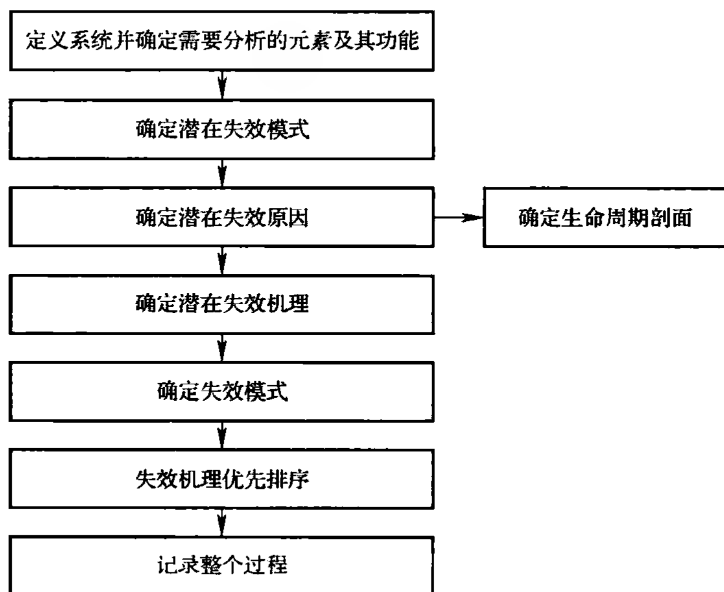


图8.2 FMMEA方法

8.2.1 系统的定义、元素和功能

FMMEA过程首先从定义将要分析的系统开始。系统是为了达到某个特定的目标，

将一些子系统或者不同层级的零部件整合在一起的组合。系统可以分成不同的子系统或层。这些子系统可能包括更多的子内容，或多个组成此子系统的部分。这些部分称为“部件（Component）”，它们形成了产品的基础结构。

为了满足分析小组的需要或为了方便起见，可以根据功能（即根据系统元素是“做”什么的）或位置（即根据系统元素“是”什么）来划分系统，或者同时根据这两者来划分（即系统中在某一位置的系统元素的功能，或系统中具有同样功能的系统元素所处的位置）。例如把汽车作为一个系统，根据功能就可以分为冷却系统、制动系统和动力系统，根据位置可以划分为发动机舱、乘客舱和仪表盘或控制板。在印制电路板系统中，根据位置划分为封装、镀通孔（PTH）、金属喷镀和电子板。这样就可以对每一个元素进行分析了。

8.2.2 潜在失效模式

失效模式（Failure Mode）是观测到即将要发生的失效的影响结果［SAE Standard, 2002］。失效模式也可以定义为一个元件、子系统或系统不能达到或完成其既定功能的情况。

对于所有已确定的元素，为每一个元素列出所有可能的失效模式。例如在一个焊接连接中，潜在失效模式是开路或焊接阻力的间歇性变化，后者可能会以互联的形式影响此连接的功能。在缺少可能发生失效模式信息的情况下，可以使用应力数值分析、加速失效试验（例如HALT）、以往经验和工程判断来确定潜在失效模式。潜在失效模式可能是高级别子系统、系统中失效模式的原因，也可能是低级别元件失效模式的结果。

8.2.3 潜在失效原因

失效原因（Failure Cause）是在设计、制造或使用过程中导致失效模式的事件［IEEE Standard 1413.1-2002, 2003］。此阶段为每一种失效模式列出失效的所有出现方式。在设计、制造、存储、运输或使用条件中查找可能引起失效的根本原因，这样就能确定失效原因。对于一个给定的系统元素，潜在失效原因的相关知识有助于确定激发失效模式的潜在失效机理。例如假定在一个汽车的发动机罩下，印制电路板的电子元件有一个失效的焊接连接。焊接的失效模式，如开路和焊接阻力的间歇性变化，其潜在原因是温度循环、随机振动和（或）冲击影响等条件下的疲劳。

8.2.4 潜在失效机理

失效机理（Failure Mechanism）是一个引发失效的物理、电、化学和机械应力以特定形式结合的过程［Hu et al, 1993］。为潜在失效模式和失效原因选择对应的失效机理，可以确定潜在失效机理。电子材料失效机理和物理损伤模型在可靠电子产品设计中的应用研究包括所有和电相关的磨损和过应力失效，在相关文献中可以找到这些内容［Dasgupta and Pecht, 1991; JEDEC, 2003］。

因此，可以把失效机理分类为过应力（Overstress）和磨损（Wearout）机理。过应力失效是由于单一载荷（应力）作用条件下引起的失效。与之对应，磨损失效则是由于累计载荷（应力）引起的失效［IEEE Standard 1413.1-2002, 2003］。例如在焊接连接的例子中，受振动和冲击导致焊点开路和短路的潜在失效机理分别是疲劳和过应力

冲击。

8.2.5 失效模型

失效模型使用合适的应力和损伤分析方法评估失效敏感度。通过评估失效时间和给定几何形状、材料结构、环境和运行条件下的失效可能性，可以估计失效敏感度。例如在焊接连接疲劳的例子中，Dasgupta 失效模型 [Dasgupta et al, 1992] 和 Coffin-Manson 失效模型 [Foucher et al, 2002] 可用来分析温度循环的应力和损伤。

在已定义应力条件下，过应力失效机理模型使用应力分析来估计失效的可能性。过应力模型最简单的公式化表示是：诱导应力（Induced Stress）与必须承载应力的材料强度的比值。在已定义应力条件下，磨损机理同时使用应力和损伤分析计算诱发失效所需的时间。在磨损失效中，损伤会在一定时间内累积，直至材料不能再承受外加载荷。因此，对于多种条件结合的情况，必须要找出合适的方法，以估计失效时间。有时候，由于单个载荷条件引发的损伤可以单独分析，但失效评估结果可能要以累加的形式结合起来 [Guidelines for Failure Mode and Effects Analysis for Automotive, Aerospace, and General Manufacturing Industries, 2003]。

限制失效时间评估的因素可能包括：

① 用于量化系统失效时间模型的可用性、精度。

② 结合多失效模式结果和单失效模式结果的能力。

③ 相同模型结合多应力条件的能力 [IEEE Standard 1413.1-2002, 2003]。

如果没有可用的失效模型，可以根据经验模型来选择合适的参数以进行监测。经验模型来自于以前的现场失效数据或者从加速试验中得到的数据。

8.2.6 生命周期剖面

生命周期剖面（Life-Cycle Profile）包括环境条件，如温度、湿度、压力、振动或冲击、化学环境、辐射、污染物，以及由于运行条件引起的载荷，如电流、电压和功率 [SAE, 1978]。产品的生命周期环境包括装配、存储、运输以及产品的使用条件，还包括这些条件的严重度（Severity）和持续时间。生命周期条件的相关信息可用来消除那些在给定运行条件下不会发生的失效模式。

在缺少现场数据的情况下，可从环境手册中获得产品使用条件的信息，也可以使用相似环境的监测数据。理想的方法是获取这些数据，并在实际的应用中对它们进行处理。在相同或相似产品生命周期环境中记录的数据可以作为 FMMEA 过程的输入内容。一些组织收集、记录，并以手册的形式发布了这些数据。一些设计师和工程师根据自己兴趣为特定市场开发产品，这些所发布的手册为他们提供了指导。这些手册能提供最接近于产品运行时的环境条件，还特别提供了环境变量的总值，但没有涵盖所有的生命周期条件。例如对于一般汽车应用的生命周期环境和运行条件，可以从 SAE 手册 [SAE, 1978] 查询到。然而，对于特定的应用，还需要获得更为详细的信息。

8.2.7 失效机理的优先排序

理想状况下，必须为产品设计和分析考虑所有失效机理以及这些机理之间的相互影响。在一个产品的生命周期中，不同的环境、各种应力级别下的运行参数都可能激活一

些失效机理，但是在一般情况下，只有一少部分的运行、环境参数和失效机理是大部分失效的原因。高优先度失效机理是那些使得产品在预期寿命之前出现失效的机理，这些机理出现在正常运行、产品应用环境的条件下。通过对所有潜在失效机理排序，可以确定高优先度失效机理，这样可以有效地利用资源。图8.3中显示了失效机理优先排序的方法。

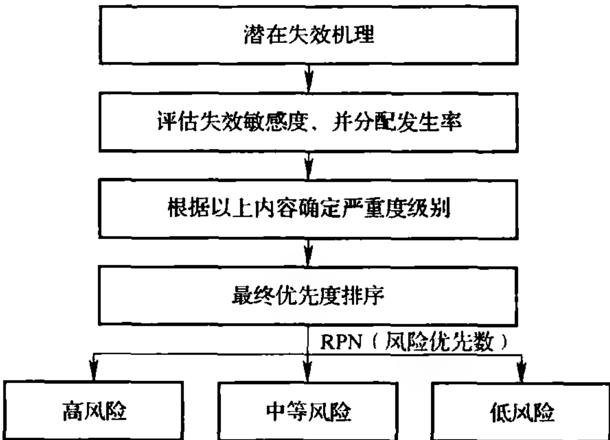


图8.3 失效机理优先度排序

首先，要为所有潜在失效机理优先度排序建立环境和运行条件。如果由某个特定的运行、环境条件没有产生载荷或者所产生的载荷级别很模糊，那么，对于仅仅由这些环境和运行条件决定的失效机理，分配给它们“低”风险级别，后面就不再考虑它。

对于初始排序后剩余的所有失效机理，使用前面确定的失效模型（当模型可用时）计算这些失效机理的敏感度。对于过应力机理，通过应力分析决定在给定环境和运行条件下，失效是否会得到暴露，这样就能计算它的敏感度了。对于磨损机理，在给定环境和运行条件下，决定其失效时间就可以得到其敏感度。为了决定所有磨损失效产生的综合影响，还需要计算所有同时发生的磨损机理的失效时间。在没有可用的失效模型时，基于以往经验、生产制造数据或手册进行计算。

在计算了失效敏感度之后，把系统在给定环境和运行条件下的失效发生率分配给失效机理。发生率描述的是某失效机理将引发失效的频率。对于激发失效的过应力失效机理，赋予它高发生率“5”（经常发生）；如果没有激发过应力失效，则给予低发生率“1”（几乎不可能发生）。对于磨损失效机理，要对其单个失效时间、所有失效时间、预期产品寿命、以往经历和工程决策进行标杆管理，这样才能分配它的发生率级别。表8.1是失效机理的发生率。

表8.1 失效机理的发生率

发生率	级别	标准
经常发生	5	过应力失效或很低的TTF
有可能发生	4	低TTF
偶然发生	3	中TTF
较少发生	2	高TTF
几乎不发生	1	非常高的TTF或没有过应力失效

发生率“经常”指失效时间（Time To Failure，TTF）很低的失效机理，或过应力失效。在使用条件下，过应力失效几乎不可避免。“有可能发生”指拥有低TTF的失效机理，“偶然发生”指拥有中等TTF的失效机理，“几乎不可能发生”分配给那些拥有

非常高 TTF 的失效或那些不产生任何失效的过应力失效机理。

为了对失效影响进行量化度量，要为每个失效机理分配一个严重度等级。严重度是失效产生影响的严重程度。首先在所分析级别评估失效的影响，然后在更高的级别、子系统级别，最后在系统级别评估失效影响［SAE Standard，2002］。安全问题和失效机理对最终系统的影响可作为分配严重度级别的主要标准。在严重度分级过程中，可能要把最严重的程度分配给正在分析的失效机理。以往的经验和工程决策可能也有助于分配严重度级别。表 8.2 中列出了所有的严重度级别。

表 8.2　严重度级别

严　重　度	级　　别	标　　准
非常高或灾难性的	5	系统失效或安全相关的灾难性失效
高	4	丧失功能
中等或显著的	3	性能逐步下降
低或较小	2	在性能下降情况下，系统可以运行
非常低或没有	1	轻微干扰

①“非常高或灾难性的”严重度级别：失效可能会引起使用者丧命，或对产品造成无法弥补的损害。

②“高”严重度级别：失效可能会引发使用者严重受伤，或产品功能的丢失。

③“中等或显著的”严重度级别：失效可能会对使用者造成较小创伤，或在可靠性丧失的过程中，随着时间的推移，产品性能有所下降。

④“低或较小”级别：失效可能不会引起任何用户受伤，或导致产品性能下降。

⑤“非常低或没有”级别：不会引起任何受伤，且对产品没有影响，在最佳情况下，只有轻微干扰。

根据每个已确定失效机理的严重度和发生率，可以计算风险优先数（Risk Priority Number，RPN）。RPN 是严重度（Sev）和发生率（Occ）的乘积。优先度排序的最后一个步骤可参考表 8.3 中给出的风险矩阵，基于 RPN 可把失效机理分成三个风险等级。分级结果可能会因为产品类型、使用情况和使用者或制造商的经营目标的不同而有所差异。

表 8.3　风险矩阵

严　重　度	发　生　率				
	5 经常发生	4 有可能发生	3 偶然发生	2 很少发生	1 几乎不发生
5 非常高或灾难性的	25 高风险	20 高风险	15 高风险	10 高风险	5 中等风险
4 高	20 高风险	16 高风险	12 高风险	8 中等风险	4 低风险

（续）

严重度	发生率				
	5 经常发生	4 有可能发生	3 偶然发生	2 很少发生	1 几乎不发生
3 中等或显著	15 高风险	12 高风险	9 中等风险	6 低风险	3 低风险
2 低或较小	10 高风险	8 中等风险	6 低风险	4 低风险	2 低风险
1 非常低或没有	5 中等风险	4 低风险	3 低风险	2 低风险	1 低风险

8.2.8 文件编制

FMMEA过程涉及到文件编制，它包括基于FMMEA要考虑的行为和采取的行动。对于已制造的产品，在产品开发和测试过程中发生的失效，要对其进行根源分析，文件编制可能就存在于根源分析的记录中。文件编制包含了历史和经验教训，这些内容为将来的产品FMMEA过程提供了框架。在采取纠正措施之后，也必须维护、更新关于FMMEA的文档，以便为将来的分析产生新的高优先度失效机理列表。

8.3 案例研究

我们以汽车中的一个简单印制电路板（PCB）组件为例，来说明FMMEA过程。PCB组件安装在1997年的Toyota四驱车的发动机舱的四个角落里，此组件包括一个表面镀铜的FR-4 PCB、镀通孔（PTH），还有八个表面贴装电感器（Surface Mount Inductor），它是用63Sn-37Pb焊料焊接在焊盘上的。电感器通过PCB表面镀铜连接到PTH。PTH由焊料填充，事件监测电路通过多个PTH串联到所有电感器，用以评估失效。该组件失效的定义是：产生崩溃，或没有电流通过事件探测器。

对于列出的所有元素，确定其相应的功能和潜在失效模式。表8.4为每个元素列出了所有可能失效模式的物理位置。例如焊接连接的潜在失效模式是断开和焊接阻力的间歇性变化。

表8.4 案例研究的FMMEA工作表

元素	潜在失效模式	潜在失效原因	潜在失效机理	机理类型	失效模型	失效敏感度	发生率	严重度	风险
PTH	PTH中的电气开路	温度循环	疲劳	磨损	CALCE PTH柱热疲劳①	>10年	很少发生	非常低	低
金属喷镀	电路短路/断开	高温	电迁移	磨损	Black②	>10年	很少发生	非常高	中等
	金属喷镀痕迹中的阻力变化	高相对湿度 粒子污染	腐蚀	磨损 磨损	Howard③	>10年	很少发生	非常高	中等

（续）

元素	潜在失效模式	潜在失效原因	潜在失效机理	机理类型	失效模型	失效敏感度	发生率	严重度	风险
零件（电感器）	线圈和铁心之间的短路/开路	高温	绕组绝缘的磨损	磨损	无		很少发生[4]	非常高	中等
布线	开路/电阻的间歇性变化	温度循环	疲劳	磨损	Coffin-Manson[5]	170 天	经常发生	非常高	高
		随机振动		磨损	Steinberg[6]	43 天	经常发生	非常高	高
		猛然撞击	冲击	过应力	Steinberg[6]	无失效	几乎不可能	非常高	中等
PCB	PTH 间的短路	高相对湿度	导电细丝生成	磨损	Rudra et al[7]	4.6 年	偶然发生	非常低	低
	裂缝/断裂	随机振动	疲劳	磨损	Basquin[6]	>10 年	很少发生	非常高	中等
		猛然撞击	冲击	过应力	Steinberg[6]	无失效	几乎不可能	非常高	中等
	聚合物强度损失磨损	高温	玻璃相变	过应力	无	无失效	几乎不可能	非常高	中等
	开路	介质材料间的高电压放电	电气过应力/静电放电	过应力	无	在第一级优先度排序已消除	低		
	过度的噪声	接近高电流或磁源	电磁干扰（EMI）	过应力	无	在第一级优先度排序已消除	低		
焊盘	翘起/断裂	温度循环/随机振动	疲劳	磨损	无		很少发生	非常高	中等
		猛然撞击	冲击	过应力			几乎不发生	非常高	中等

① Bhandarkar, S. M. et al. 1992. Transactions of the ASME—Journal of Electronic Packaging 114: 8-13。

② Black J. R. 1983. IEEE Proceedings of International Reliability Physics Symposium 142-149。

③ Howard, R. T. 1981. IEEE Transactions on CHMT 4 (4): 520-525。

④ Based on failure rate data of inductors in Telcordia. (From Telcordia Technologies. May 2001. Special Report SR-332: Reliability prediction procedure for electronic equipment, Issue 1, Telcordia Customer Service, Piscataway, NJ.)。

⑤ Foucher, B. et al. 2002. Microelectronics Reliability 42 (8): 1155-1162。

⑥ Steinberg, D. S. 1988. Vibration analysis for electronic equipment, 2nd ed. New York: John Wiley & Sons。

⑦ Rudra, A. B. et al. 1995. Circuit World 22 (1): 67-70。

为简单起见，也为了明确演示论证的目的，假设测试装置、配电板及其零件都不存在缺陷。如果在制造后进行了合适的筛选，此假设是可以成立的。另外，还必须假设组件在制造后没有受到损伤，然后为表 8.4 中的失效模式确定潜在的失效原因。例如对于焊点来说，开路和焊接阻力间歇性变化的潜在失效原因是温度循环、随机振动，或由车

体碰撞产生的突然冲击。

根据分配到失效模式的潜在失效原因，可以确定相应的失效机理。表8.4列出了已确定的失效原因的失效机理。例如对于焊点的开路和焊接阻力间歇性变化，引起失效的机理是焊点的疲劳和断裂。

对于列出的每一个失效机理，从文献中为其确定合适的失效模型。产品尺寸和几何特征相关的信息可以从设计规格、电路板的设计图样和零件制造数据手册中得到。表8.4为列出的失效机理给出了失效模型。例如对于焊接疲劳，使用Coffin-Manson [Steinberg，1988] 失效模式为温度循环进行应力和损伤分析。

组件由一个3V的电池提供电源，它独立于汽车电气系统之外。经确认，它对组件没有大电流、电压、磁或辐射影响。因为没有华盛顿区域的制造厂家数据，对于汽车发动机舱环境下的温度、振动和湿度条件，可以从SAE环境手册 [SAE，1978] 中获取相关数据。所列出的汽车发动机舱的最高温度为121℃ [SAE，1978]。假设汽车每天在华盛顿地区的两段相等距离中平均运行3h，最高冲击级别为45G/3ms，38℃时的最大相对湿度为98% [Society of Automotive Engineers，1978]。在案例研究期间，华盛顿地区平均每天的最高和最低温度分别为127℉㊀和16℃。

为每个元素确定了所有潜在失效模式、原因、机理和模型之后，可根据生命周期环境、运行条件对失效机理进行初始的优先度排序。对于此处的测试过程，把汽车发动机舱内由电气过应力（EOS）、静电放电（ESD）驱动的失效排除在外，因为没有主动元器件（Active Device）存在，且电池的电压较低。电磁干扰（EMI）也不在预期之中，因为电路功能对瞬时电压不敏感。因此，EOS、ESD和EMI都被分配成低风险级别。

用calcePWA（Maryland大学的失效物理可靠性软件）可以计算磨损失效机理的失效时间。根据特定磨损失效的失效时间与所有磨损失效机理的失效时间总和的比值，可以分配给定磨损失效的发生率。根据从文献 [Telcordia，2001] 中获取的失效率数据，分配电感器的失效发生率。根据以前与焊盘磨损相关的知识，分配给它的发生率为“很少发生”。

使用calcePWA对冲击级别45 G/3ms进行评估，其结果是它不会给布线和电路板带来失效。因此，分配给它的发生率为“几乎不可能发生”。因为我们没有预期电路板和布线会发生过应力冲击失效，所以假定焊盘处没有失效发生。过应力失效在焊盘上的发生率为“几乎不可能发生”。焊盘的玻璃相变温度为150℃。因为发动机舱焊盘的最高温度只有121℃ [SAE，1978]，所以玻璃相变将不会发生，分配给它的发生率为“几乎不可能发生”。

短路或开路的PTH对于电路没有任何影响，因为它仅用作电感器的终端。分配给它的发生率为“非常低”。对于所有其他元素，任何给定失效模式都会引起电路功能的受损。因此，其他元素的严重度级别为“非常高”。

㊀ 1℉ = (1℃ ×9/5) +32。

表8.4给出了失效机理优先度排序的结果和风险评价。所有已分析的失效机理中，在焊点处，由温度循环和振动引起的疲劳是唯一拥有高风险级别的失效机理。因为它们是高风险失效机理，因此给予其高优先度。

FMEA要确定组件的所有元素、它们的功能、潜在失效模式和失效原因，然后还要为每种失效模式确定失效影响。例如在焊点中，开路将导致没有电流在测试装置通过。然后，FMEA还要确定每个失效模式的严重度、发生率和可检测概率，例如在焊点的开路失效模式中，根据以往经验和工程决策，为每一个严重度、发生率和检测率设定1～10的级别。严重度、发生率和检测率的乘积将用于计算RPN。其他失效模式RPN的计算方法与此类似。然后，基于RPN对所有失效模式进行优先度排序。与FMMEA不同的是，它使用失效机理和模式，以及所有失效机理的综合影响来量化计算发生率。焊点处的发生率和严重度将用来为每一个失效机理的优先度排序分配风险级别。

8.4 总结

FMMEA允许设计团队考虑可用的失效机理科学知识，并把这些知识与FMEA模板的系统特征和“可靠性设计”的理念、知识结合起来。FMEA过程中的优先度排序也适用于FMMEA，它用来确定在产品生命周期中可能引发失效的机理。

FMMEA与FMEA在一些方面有所不同。FMEA单独考查潜在失效模式，然后把它与未考虑的共存失效原因产生的影响结合起来。而在FMMEA中，则要考虑同时发生的失效机理的影响。FMEA要激发并检测失效，以更新和计算RPN，它不能用于持续检测随着时间推移性能下降的情况。相比之下，FMMEA不需要激发和检测失效，也不会出现与检测估计相关的不确定性因素。在FMEA的量化阶段，不需要使用环境和运行条件。最好用FMEA来消除某些失效模式。

在FMMEA的机理优先度排序中，使用环境和运行条件的应力级别确定高优先度的机理。必须在设计阶段考虑这些机理，并要对其加以控制。这种FMMEA中的优先度排序方法克服了FMEA中RPN优先度排序的缺点，后者提供了不正确的粒度感。因此，使用FMMEA可以提供更多产品可靠性量化信息。FMMEA比FMEA提供了更多的改进可靠性的机会，因为它在分析过程中考虑了环境和运行条件的特定失效机理和应力级别。

企业使用FMMEA会得到一定的益处。它提供了具体的应力条件信息，这样，验收和质量鉴定试验就能产生有用的结果。在产品开发阶段使用失效模型，可以对技术升级提议进行适当的“假设”分析。FMMEA还有助于优化一些设计和开发步骤，只要利用失效机理和模型的知识，就能使这些步骤有所改进。这些步骤包括虚拟鉴定、加速试验、根源分析、寿命消耗监测和预测。这些措施带来的所有技术和经济效益都得益于FMMEA的使用。

参考文献

Bhandarkar, S. M. et al. 1992. Influence of selected design variables on thermomechanical stress distributions in plated through hole structures. Transactions of the ASME— Journal of Electronic Packaging 114: 8-13.

Black, J. R. 1983. Physics of electromigration. IEEE Proceedings of International ReliabilityPhysics Symposium 142-149, Phoenix, AZ.

Bowles, J. B. 2003. Fundamentals of failure modes and effects analysis. Tutorial Notes Annual Reliability and Maintainability Symposium, Tampa, FL.

Bowles, J. B., and R. D. Bonnell. 1998. Failure modes, effects and criticality analysis—What is it and how to use it. Tutorial Notes Annual Reliability and Maintainability Symposium, Anaheim, CA.

Dasgupta, A., C. Oyan, D. Barker, and M. Pecht. 1992. Solder creep-fatigue analysis by an energy-partitioning approach. ASME Transactions, Journal of Electronic Packaging 114 (2): 152-160.

Dasgupta, A., and M. Pecht. 1991. Material failure mechanisms and damage models. IEEE Transactions on Reliability 40 (5): 531-536.

Foucher, B., J. Boullie, B. Meslet, and D. Das. 2002. A review of reliability predictions meth-ods for electronic devices. Microelectronics Reliability 42 (8): 1155-1162.

Franceschini, F., and Galetto, M. 2001. A new approach for evaluation of risk priories of failure modes in FMEA. International Journal of Production Research 39 (13): 2991-3002.

Guidelines for failure mode and effects analysis for automotive, aerospace, and general manu-facturing industries. 2003. Ontario, Canada: Dyadem Press.

Howard, R. T. 1981. Electrochemical model for corrosion of conductors on ceramic substrates. IEEE Transactions on CHMT 4 (4): 520-525.

Hu, J., D. Barker, A. Dasgupta, and A. Arora. 1993. Role of failure-mechanism identification in accelerated testing. Journal of the IES 36 (4): 39-45.

IEEE Standard 1413.1-2002. 2003. IEEE guide for selecting and using reliability predictions based on IEEE 1413.

JEDEC Publication JEP 122-B. August 2003. JEDEC Publication JEP 122-B. Failure Mecha-nisms and models for semiconductor devices.

JEDEC Publication JEP 148. April 2004. JEDEC Publication JEP 148. Reliability Qualification of semiconductor devices based on physics-of-failure risk and opportunity assessment.

Kara-Zaitri, C., A. Z. Keller, and P. V. Fleming. 1992. A smart failure mode and effect analysis package. Annual Reliability and Maintainability Symposium Proceedings, 414-421.

Rudra, A. B. et al. 1995. Electrochemical migration in multichip modules. Circuit World 22

(1): 67-70.

SAE (Society of Automotive Engineers). Rev. November 1978. Recommended environmental practices for electronic equipment design, SAE J1211.

SAE Standard. August 2002. SAE J1739 Potential failure mode and effects analysis in design (design FMEA) and potential failure mode and effects analysis in manufacturing and assembly processes (process FMEA) and effects analysis for machinery (machinery FMEA). Steinberg, D. S. 1988. Vibration analysis for electronic equipment, 2nd ed. New York: John Wiley & Sons.

Telcordia Technologies. 2001. Special Report SR-332: Reliability prediction procedure for electronic equipment, issue 1, Telcordia Customer Service, Piscataway, NJ.

University of Maryland. A physics-of-failure-based virtual reliability assessment tool developed by CALCE, The University of Maryland.

第9章　可靠性设计

9.1　引言

企业必须在产品开发过程中实施一些措施，以保证产品的可靠性。这些措施会通过零件（材料）选择、产品设计、制造、装配、运输、装卸、运行、维护和修理等过程影响产品的可靠性。本章将介绍以下内容：

① 根据目标生命周期应用条件和预期产品性能等因素，制定切实可行的产品可靠性要求。制定这些要求时，必须考虑客户需求，以及制造商满足这些要求的能力。

② 评估相关的制造、装配、存储、装卸、运输、运行和维修条件，以此来定义产品生命周期条件。

③ 确保供应链参与者有能力提供符合要求的零件（材料）和满足最终可靠性目标所需的服务。

④ 选择有质量保证的零件（材料），这些零件要能在应用中实现产品的预期性能和可靠性。

⑤ 确定可能发生失效产品的潜在失效模式、失效位置和失效机理。

⑥ 在设计加工能力（也就是制造和装配中可控制的质量级别）时，考虑潜在失效模式、失效位置以及从失效物理分析和生命周期剖面中获取的失效机理。

⑦ 对产品进行鉴定，以在预期生命周期条件下验证产品的可靠性。鉴定包括所有能保证标称设计和制造规范将满足或超过可靠性目标的活动。

⑧ 所有制造和装配工艺必须能产生出在设计所要求的统计工艺窗口内的产品。材料特性和制造工艺的变异性将影响产品的可靠性。因此，整个过程必须是经过鉴定、可度量以及可监控的。

⑨ 利用闭环根源监测程序管理产品的生命周期运行情况。

9.2　产品需求和约束

一个产品的开发、改进或升级有着各种各样的原因，例如企业想占领某个潜在市场，或是开辟新的市场。在某些情况下，企业需要开发新的产品，以此在主要市场保持竞争力或保持目前的市场份额和客户忠诚度。另外一些情况是，企业想满足某个关键客户，或者要用新技术或方法来证明自己在某些方面的经验，或者改进已存在产品的维护性。此外，已存在产品的升级通常是为了减少生命周期费用。

为了生产可靠的产品，供应商需要通过供应链和客户相互协作。有文献［IEEE

1332, 1998] 从可靠性目标的三个方面阐述了这种协作关系。首先，与客户一起工作的供应商需要决策并理解客户要求和产品需求，这样才能制定全面的设计规范；其次，供应商需要组织、实施一系列的工程活动，才能生产出能满足客户要求、具有一定可靠性的产品；最后，供应商需要展开一些活动，以保证满足客户对可靠性和产品的要求。

首先把客户需求写入需求文档中，然后对它们进行优先级排序。参与排序和最终决策的人员会因企业组织和产品的不同而有所差异。例如对于安全关键产品（Safety-Critical Product）来说，与安全、可靠性相关的人员和法定代表人都需要对整个过程进行监督。

一旦确定了客户需求，产品工程函数要以规范的形式对需求做出反馈。规范规定了必须满足的需求，满足需求所需的日程计划、参与工作的人员、潜在的风险等。最初制定的规范和需求文档的差异将变成权衡分析的主要内容。

只要定义了产品需求并启动了设计过程，就需要对产品需求与实际产品设计情况的符合程度进行评估。随着产品设计变得细节化，追踪产品特征与最初产品需求的关系将变得越来越重要。要记录所做出修改的根本原因。完整的需求追踪能大幅度减小未来产品重新设计的成本。通过技术监测，制定重新设计或设计更新计划，使用路线图以确保企业能够及时将新产品或重新设计产品推向市场，这是保持客户基础、确保持续利润的有效途径。

9.3 产品的生命周期条件

产品的生命周期条件影响着关于产品设计和开发、材料和零件选择、鉴定、产品安全性、质量保证和产品保障（也就是维修）的决策等。产品生命周期的阶段包括制造和装配、试验、返修、仓储、运输、装卸、运行[⊖]（运行模式、启动/停止周期等）、修理和维修等。

在生命周期的每个阶段，产品将经历各种环境和使用载荷。生命周期载荷包括且不局限于热（稳定状况温度、温度范围、温度循环、温度梯度）、机械（压力级别、压力梯度、振动、冲击载荷、噪声级别）、化学（恶劣或惰性环境、臭氧、污染湿度级别、污染物、燃料泄漏）、辐射（电磁干扰和强度）和电载荷条件（功率、功率骤增、电流、电压、电压尖峰）。因此，产品退化的程度和速率、可靠性取决于这些载荷的性质、强度，以及产品暴露在这些载荷下的持续时间等。

通常，整个可靠性设计过程的生命周期载荷定义和描述是一个不确定因素，其原因是产品可能经历完全不同的应用条件。这些应用条件取决于运行定位、产品的使用或非使用剖面、使用持续时间、维修和服务条件。桌面计算机就是这样一个典型的例子，所有的桌面计算机都是为家庭或办公环境而设计的，但是每台计算机的运行剖面可能完全不同，这取决于用户的行为。一些用户可能会在每次使用后关掉计算机，而有些人则可

⊖ 有时也把运行条件称为生命周期应用条件。

能在每天的最后时刻才会关闭计算机，也有些人则会一直让计算机保持开机状态。另外，某个用户可能把计算机放在能照到太阳的窗户旁边，另一个用户可能把计算机放在空调旁边。因此，每个产品经历的温度剖面是不同的，它因热载荷引起的性能下降情况也将有所不同。

估计产品生命周期载荷的方法有四种：市场研究和行业标准、相似性分析、现场试验和服务记录、原地监测。下面将介绍这些方法。

对于不同应用情况可能存在的环境载荷，市场研究和行业标准只能提供一个非常粗糙、且通常并不精确的估计。可以基于工业类型（如军事、消费、远程通信、自动化和商业航空电子），对来自于市场调查和行业标准的环境剖面进行分类。

当拥有同类产品足够多的使用现场历史资料时，相似性分析可以作为估计环境载荷的技术。在把已存在产品的数据应用到提议设计之前，需要评审设计产品和对比产品应用情况的差异。例如在商业洗衣房内，洗衣机内部的电子设备将可能会经历广泛分布的载荷和使用条件（由于存在大量的用户），与家用洗衣机相比，它的使用率更高。另外一个例子是，经研究发现，一些亚洲人除了用洗碗机清洗餐饮用具外，还用它清洗蔬菜。这些洗碗机比那些只用来洗餐具的洗碗机的使用频率就要高一些。

现场试验记录用于估计产品经历的环境剖面，它产生的数据取决于试验的持续时间和条件，也可以用它来预测估计实际环境情况。服务记录提供了对产品进行维修、更换或所提供服务的信息。一些生命周期环境和使用条件会引发维修活动或产品失效，服务记录的数据能提供关于这些环境和使用条件的资料。

对于产品在生命周期中所经历的环境和使用条件，还可以对其进行原地监测［Vichare et al，2004］。通常用扩展安装或集成到产品的传感器来收集这些数据，传感器要有远距离试验系统的支持。可以对不同用户使用的产品进行监测，用得到的数据推断载荷分布，这些产品最好来自不同的地理位置。在充足的时间段内收集数据，这样才能对载荷及其在此时间段内的变化进行估计。原地监测提供了最精确的载荷历史数据，对于可靠性设计（Design For Reliability，DFR）和产品可靠性评价而言，它是最具使用价值的方法。

9.4 可靠性能力

通常，要根据一些不能明确表述可靠性的因素选择供应链，如技术能力、生产规模、地理位置、保障设施，还有财务和合同因素等。选择过程需要考虑供应商完成制造、试验中可靠性目标的情况，对最终产品在生命周期中可靠性的保障能力，是否具有重要的竞争优势等。

对于致力于提高最终产品可靠性的企业组织来说，可靠性能力是对其所实施措施的度量，它用来度量这些措施对客户可靠性需求的满足情况。可靠性能力评估使用一个称为可靠性能力成熟度（Reliability Capability Maturity）的量对可靠性活动进行量化评估。从可靠性的角度来讲，成熟度意味着一个组织实施的关键可靠性措施是否是经过充分理

解的、是否受到文件编制和训练的支持、是否应用到了此组织的所有产品中、是否得到了持续的监测和改进。

9.5　零件和材料选择

零件（材料）选择和管理方法有助于企业做出一些风险指引决策，这些决策与把零件和材料组合（装配）到产品中相关。图 9.1 是零件评估流程。零件评估的主要内容包括：性能、质量、可靠性、易于装配性等。

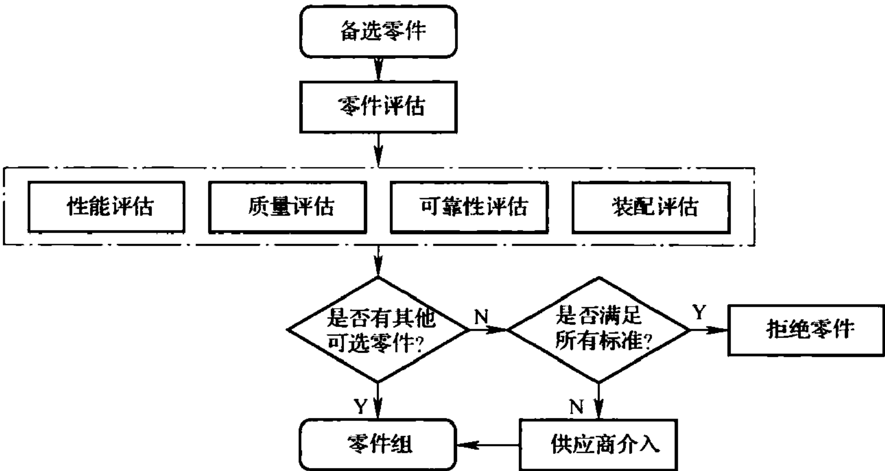


图 9.1　零件评估流程

性能评估的目的是评价零件满足产品性能需求（例如结构、力学、电、热、生物等方面的需求）的能力。一般来说，这些性能需求都有上限和下限，当零件超过界限时，它将不能实现其应有的功能。零件至少要满足数据表规范。通常把这些限定或级别称作推荐运行条件。

质量是用输出质量和工艺能力评价的。可靠性评估的结果提供了零件在目标生命周期应用中的某个特定时间段内满足性能规范的情况。通过零件鉴定和可靠性试验得到的结果，可以用来评估可靠性。

从装配的观点来看，如果零件能够与下游装配设备和过程相兼容，那么它就是可接受的。在装配过程中要遵循装配的指导原则，才能阻止零件的损坏和恶化。这些原则包括：推荐温度剖面、清洗剂、粘合剂、对湿气的灵敏度和电气保护等。随着新兴的技术出现，产品变得越来越复杂，装配指南也变得更加重要，它能保证零件和产品的目标质量和可靠性。

9.6　失效模式、机理及影响分析

失效模式是失效发生的方式，也就是说，产品不能实现其既定设计功能，或实现了

功能却没有达到预期目标。例如手机电话的失效模式包括：某个按键不能实现数字的输入，麦克风不能采集到用户的声音。

有时，要特别加重某些失效模式，这样产品用户才会意识到存在的问题。例如把有不良气味的物质添加到天然气，以提示发生了泄漏。另一个例子是汽车中制动垫磨损时发出的研磨噪声。

失效机理是引发失效的物理、电、热和机械应力的特定结合过程。例如断裂、疲劳和腐蚀都是失效机理。

实施失效模式、机理及影响分析（Failure Modes, Mechanisms, and Effects Analysis, FMMEA）的目的是为产品所有潜在失效模式确定潜在失效机理和模型，然后为高效的产品开发提供失效机理的优先级排序。实施FMMEA的基础是了解产品需求和产品物理特征（及其在生产过程中的变化）的关系；理解产品材料和载荷（在应用条件下的应力）的相互影响；在使用条件下，理解材料和载荷对产品失效敏感性的影响等。

9.7 失效物理

只要根据FMMEA确定了零件（材料）、载荷条件和可能的失效风险，基于失效物理模型的设计指南就有助于决策设计权衡，也可用它来确定试验、筛选和降额[⊖]（De-Rating）因素。可以把基于失效物理模型的试验规划为可量化度量的，这样就能检测到意外的瑕疵，检测到制造或维修的问题所在。在不影响产品设计寿命的情况下，规划筛选过程以暴露“弱”产品中的失效。决策安全因素的降额能为关键失效机理降低应力。

9.7.1 应力裕度

产品的设计需要能在推荐运行范围（规范界限）之外的裕度（设计裕度）内满意地运行。这些裕度范围一定要包括在采购需求和规范中。

图9.2图形化地描述了产品载荷（应力）界限和裕度的层级。规范界限由制造商制定，用来限定用户的使用条件。设计裕度与载荷（应力）条件相对应，是产品设计能无现场失效存活的条件。也就是说，运行裕度是一些将可能导致永久（过应力）失效的预期载荷（应力）。

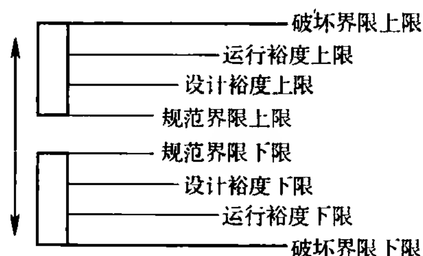


图9.2 载荷（应力）界限和裕度

产品参数变化产生的影响可用统计分析和最坏情况分析进行评估。在统计分析中，要建立产品的输出特征和参数的函数关系；在最坏情况分析中，要根据寿命结束时的性能值来评价产品的输出。

⊖ 降额是把零件置于低于它所能承受的电或机械应力下，以改进预期寿命的措施。

9.7.2 失效机理的模型分析

失效机理的模型分析基于计算机辅助模拟进行。它有助于确定生命周期载荷下产品的主要失效机理，并对其进行分级；有助于为给定加速试验参数决策加速因子；有助于决策已确定失效机理相应的失效时间。

每个失效模型都包括一个载荷分析模型（Load Analysis Model）和一个损伤评估模型（Damage Assessment Model），其输出结果是根据失效时间确定的不同失效机理的分级。根据某材料对外施载荷的影响，载荷模型可得出产品结构和损伤模型。失效机理的模型分析可用来优化产品设计，其原则是产品的最小失效时间大于预期寿命。虽然从失效机理模型分析获得的数据不能完全取代从物理试验取得的数据，但是它们暗示了预期的潜在失效模式和机理，这可以提高试验效率。

需要记住的是，模态分析结果的精度取决于过程输入的精确性，即产品的几何和材料特征、生命周期载荷、所使用的失效模型（例如失效模型中的一些常数）、域分析和离散化方法（空间和时间）。

9.7.3 降额

使用降额可以确保产品在如图 9.2 所示的预先决定的裕度内运行。降额是一种限制载荷（例如热、电和机械载荷）的措施，它可以改进产品可靠性。降额可以为设计师在不可预见的异常情况下提供保护（例如瞬态负载和电涌），例如电子零件制造商通常为电源电压、输出电流、功率消耗、结点温度和频率制定限制条件。产品设计小组可以决定降额，以确保特定载荷（如温度）的运行条件在额定水平之下。当失效机理是磨损类型时，载荷的降低能延长产品的使用运行寿命。当失效机理是过应力类型时，降额也能通过设定安全裕度提供安全运行的条件。

就像“降额”一词本身所暗示的含义一样，此方法包括两个步骤：首先决定载荷的“标称”值，然后分配给它一个较低的值。降额过程提供的安全裕度是实际最大可允许外施载荷和产品的标称限制之间的差值。

为了确保有效性，降额的目标设定必须合理，一定要根据相关失效机理建立关键载荷参数。一旦为关键失效机理确定了失效模型（例如 FMMEA），就可以在给定载荷下确定降额对产品有效可靠性的影响。我们的目标是为产品决定“安全”运行的范围，然后在此范围内运行产品。

9.7.4 保护结构

建立保护结构的目的是在初始失效或故障发生后，能采取一些行动以阻止额外或二次失效。保护技术包括使用熔丝、断路器、自感应结构和用于修正参数偏离的调整结构。

在以安全性为目标的设计中，通常需要结合一些方法以阻止产品的失效或阻止失效时引发进一步的损伤。熔丝和断路器是很好的例子，它们用来自动检测过电流或过电压尖峰，然后为电子产品切断电源。相似的例子是恒温器，它可以用来感知关键温度限值条件，然后为产品切断电源，直到温度恢复正常。自检电路也能用来感知反常条件，然后使其恢复正常，或起动弥补故障的电路。

一些情况下，产品会发生部分失效，但要允许其他部分运行，虽然这样可能会导致性能降低，但至少产品不会完全停止运行。例如某电路在另一控制产品的死区（Dead Band）内提供精确微调调整，当它在紧急条件下失效关闭时，只单独使用死区控制电路或许也能达到可接受的性能要求。

在设计保护结构时，必须要考虑维修对它的影响。例如某保护电路的熔丝被取掉，接下来将面对的问题是：产品重新通电时，将会产生什么样的影响？什么样的保护结构才适合于修理后的产品运行？在有或者没有失效安全保护结构时，需要制定并遵循什么样的维修指导原则？

9.7.5 冗余

冗余的目的是：即便产品的一个或多个零件失效，也要保证产品的成功运行。设计小组经常发现，如果没有充裕的时间采取其他措施，冗余将是改进产品可靠性最快的方法。如果可靠性需求超过了目前的工艺水平，它是最具经济效益的解决方法，或许也是唯一的方法。

典型的冗余设计是增加尺寸、重量和成本。实施不当时，它也会对可靠性造成负面影响。如果某失效原因能在同一时间影响产品所有的冗余元素，那么冗余将无法带来任何益处。当然，即便在有冗余存在的情况下，传感器、开关电路或软件的失效仍然会导致产品失效。

9.7.6 预测

根据产品状况可预测正常（按照物理性能定义）运行条件的偏离或恶化程度［Vichare et al，2004］。关于产品正常的知识可用于检测并隔离故障或失效（诊断），这样就能根据当前条件预测即将发生的失效（预测）。因此，根据实际生命周期条件决定失效的发生情况，可以开发出缓解、管理潜在失效并能对产品进行维修的程序。

可以通过以下方法预测产品设计的失效：

① 安装内置熔丝和熔断结构，在生命周期条件下，这些结构将比实际产品更快地发生失效［Mishra and Pecht，2002］。

② 检测具有失效前兆的参数，如缺陷或性能恶化［Pecht et al，2001］。

③ 检测生命周期环境和影响系统正常的运行载荷，使用失效物理模型对剩余有效寿命进行估计，可以得到一些度量数据。运行载荷还会影响处理度量数据的过程［Mishra et al，2002；Ramakrishnan and Pecht，2003］。

9.8 鉴定

在产品的使用过程中，会出现一些潜在失效，鉴定试验用来确定和评估这些失效。它需要在产品开发初期执行，对现有产品进行任何重要设计或生产变更之后，也要执行鉴定试验。

在一些情况下，产品的目标应用条件（也就是使用条件）可能是未知的。例如开发出的某组件的一个零件出售到开放性市场中，它可能会组装到多种不同类型的产品

中。此时需要执行标准鉴定试验。但是通过了这样的试验，并不意味着产品在实际目标应用中就是可靠的。因此，一般要确保目标应用中的最终产品的可靠性，仅对产品零件（材料）进行鉴定试验是不够的。

在实际运行条件下，通常没有充足的时间为产品进行完整的目标应用寿命试验。因此，常常需要采用加速（鉴定）试验的办法。

加速试验的基本思想是：在高载荷条件、短时间内，产品将出现的失效机理与在实际生命周期载荷条件、长时间内产品出现的失效机理是一样的。其目的是为获取产品可靠性信息减少总时间和成本消耗。

加速试验可分为两类：定性试验和定量试验。定性试验通常为产品施加过应力，以确定将引起过应力失效或早期磨损失效的载荷条件。这些试验适用于单载荷条件，如冲击、温度极限或电过应力等，也可用于以上载荷相互结合的情况。试验的结果包括失效模式信息，但定性试验通常不能在实验过程中估计失效时间。

定量试验的目标是磨损失效机理，这类机理是载荷条件累积的结果。从加速环境到使用环境的定量推断具有一定可信度，定量试验使得定量推断的分析成为可能。

加速寿命试验最简单的形式是连续使用加速。此方法的目标是尽可能地把产品寿命压缩到最短时间。它假定产品不会持续使用，当不使用时，产品没有载荷（应力）。假定大部分洗衣机平均每周的使用时间是 10h。如果某洗衣机连续运行，可以把加速因子[⊖]定为 $24\times7/10=16.8$。因此，如果产品的质量保证或设计寿命为 5 年，那么就需要对产品进行为期（5/16.8）年 =0.3 年，或 106 天的试验。

对于高利用率或拥有较长预期寿命的产品来说，连续使用试验不是非常有效的办法。在此情况下，加速试验用来度量产品在设定载荷（应力）下的性能表现，设定的载荷（应力）比正常遇到的情况要严重得多，这样才能保证损伤累积会在较短时间内加速进行。此类试验的目标是加速时间相关的失效机理；加速损伤累积速率以减少失效时间。根据从加速试验获得的数据可以推断产品在目标使用条件下的失效时间。

在加速试验的开始阶段，要确定失效模式、机理和影响分析（FMMEA）中所有重要的过应力和磨损失效机理。选择引发失效机理的载荷参数作为加速参数，通常把它们称作加速载荷。常见加速载荷包括热载荷，如温度、温度循环和温度变化率；化学载荷，如湿度、腐蚀、酸、溶剂和盐；电载荷，如电压或功率；机械载荷，如振动、机械载荷循环、应变循环和振动/冲击等。加速试验需要把这些载荷结合起来进行。要对综合载荷的结果进行表述，就需要对它们之间的相互影响有量化的认识。

一些加速参数会诱发由特定机理引发的失效。例如腐蚀会因温度和湿度加速；蠕变会因机械应力和温度原因而加速。此外，单载荷的加速会引发一些机理的失效，例如温度能加速众多失效机理的磨损损伤累积，如腐蚀、电化学迁移和蠕变。在不正常运行条件下，主要的失效机理可能会因为载荷的改变而失去其地位，例如高功率电子产品产生

⊖ 加速因子是指产品在正常使用情况下的寿命与在加速条件下的寿命之比。

的温度能馏出湿气；相反，在正常使用条件下隐匿的失效机理可能会在加速条件下引发设备失效。因此，在没有额外的失效机理或没有可具有代表性的物理或材料行为出现时，如果加速试验要加速实际使用环境和运行环境，就要谨慎地对其进行规划。

一旦确定了失效机理，就需要选择合适的加速载荷；决定试验步骤和载荷级别；决定试验方法，如恒定载荷加速或步增载荷加速；进行试验；解释试验数据，包括从加速试验结果推断正常运行条件。试验结果为评估产品可靠性、改进产品设计、规划质量保证和保障提供了失效信息。

9.9　制造和装配

制造和装配过程对产品的质量和可靠性有很大的影响。不适当的制造和装配会引发缺陷、瑕疵和残余应力，残余应力表现为潜在失效位置，也会在产品寿命后期增加（或提高）应力。制造过程的变异对失效的影响如图9.3所示。

在制造过程中，产品强度有时会降低，关键参数标准偏差的平均值或增加值会因此而引发早期失效。一般来讲，需要对产品进行鉴定，以保证产出产品是可靠的。此时需要进行批次间筛选，以确保装配和制造相关参数的变异在定义公差范围内。在产品到达最终客户手中之前，筛选通过暴露潜在缺陷确保了产品的质量。

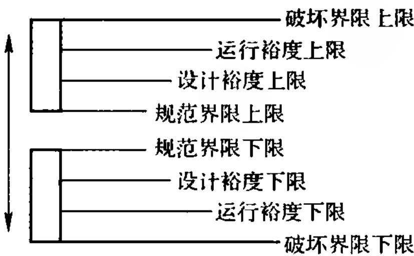


图9.3　质量对失效概率的影响

9.9.1　工艺性

设计小组必须了解材料特性和构建产品的工艺过程能力，这样才能改进可制造性、减少缺陷的发生。设计小组必须为可接受质量定义明确的阈值，并定义不合格产品。质量不合格的产品不能通过验收。

缺陷是过程的产物，在任何时间内，所有能损害或有潜力损害产品性能的产物都是缺陷。它可能出现于单个过程中，也可能是一系列过程的结果。一个过程产出的成品是那些能在后来过程或生命周期中使用产品的一部分。缺陷的来源有时会比较模糊，因为直到产品到达开发过程下游的某个点之前，可能都不会检测到某个过程中所产生的缺陷。

通常可以简化过程，以减少工艺产生缺陷的概率。随着工艺变得复杂，需要执行工艺检测和控制才能确保产出没有缺陷的产品。通常，定义工艺是否在容许界限值内的裕度被称为过程窗，它是根据工艺中要控制的独立变量、工艺对产品的影响定义的。其目标是帮助工程人员理解每个工艺变量对每个产品参数的影响，这样才能为工艺制定控制界限，也就是缺陷率开始拥有引发失效的潜力。

在定义过程窗时，必须定义每个工艺变量的上限和下限，工艺超过这些限制会产生缺陷。通过缺陷试验、缺陷原因分析，并用工艺控制消除缺陷，如通过闭环式修复行为

系统来控制工艺，一定要把制造工艺控制在过程窗之内。为工艺相关的缺陷数据汇报建立一个有效的反馈途径是关键所在。一旦完成了这样工作，也就可以确定过程窗了，它可作为一个工艺操作员的反馈系统。

一些工艺参数可能会相互影响，这样产生的缺陷不同于单独参数独立作用产生的缺陷。在这种复杂的情况下，需要设计专门的试验，用它来对各种工艺参数的相互影响进行评价。

在一些情况下，直到过程的后期才能检测到缺陷。在为产品添加了大量的价值之后，缺陷能引发产品的驳回、返工或失效。通过增加隐藏的工厂成本，这样造成的成本使得投资减少。对于所有关键工艺，需要专门通过工艺控制来消除缺陷。

9.9.2　工艺验证试验

过程验证试验通常称为筛选。它涉及对所有制造产品进行 100% 的审查，以检测或暴露缺陷。此步骤的目的是在产品使用现场之前预先找出潜在的质量问题。因此，筛选有助于减少质保的返修，有助于增加客户友好度。原则上讲，如果选择了合适的零件（材料），如果对过程进行了良好的控制，就不需要筛选过程。

一些产品会出现失效的多峰概率密度函数情况，由于使用不适当的材料、制造和装配技术的控制或运行不当，峰值会出现在产品服务寿命的早期阶段。这种早期寿命失效通常称为早期失效（Infant Mortality）。运用适当的筛选技术，可以成功地检测或暴露这些失效，消除或减少它们在使用现场的发生率。如果需要，只有在生产的早期阶段，才考虑使用筛选，也只有在产品将出现早期失效时才使用它。如果在失效概率密度函数中只有一个峰值，那么筛选就会变得无效，且执行成本很高。此外，我们不能有效地筛选由于不可预见事件，如闪电或地震引发的失效。

因为筛选是在 100% 产品的基础上进行的，所以开发出不损伤好产品的筛选机制非常重要。因此，最好的筛选机制是无损检测技术，如显微视觉检测、X 射线、声波扫描、核磁共振、电子顺磁共振等。应力筛选涉及应用载荷，它有可能在额定运行限制之上。如果应力筛选是不可避免的，那么要在加速磨损试验之间进行，因为后者比前者更可能伤害一些好产品的使用寿命。

在应力筛选过程中，如果对好产品的损伤是不可避免的，那么就需要根据失效机理模型对所筛选损伤进行量化估计，这样才能让设计小组考虑此过程中使用寿命的损失。筛选的应力级别必须适用于具体指定的产品。随着鉴定试验的进行，失效机理的量化模型有助于决策筛选参数。

应力筛选不需要模拟现场环境，在现场条件下，甚至可以使用由失效触发的失效机理模型的相似模型；反之，筛选机制需要利用最方便、有效的失效机理去模拟缺陷，这些缺陷会在使用现场以早期失效的形式出现。这就要求相关人员对产品中可能出现的缺陷有清晰的认识，并要熟悉相关的失效机理。

任何将要进行的应力筛选都需要财力和人力资源，它们来为所有失效单元决策根源和适当的修复行为。用来进行筛选的应力类型，应该由设计、制造、质量小组中适当的人员来选择。在生产过程早期，有可能需要进行应力筛选，但它会为资本、运行开销和

周期时间带来一些负面影响。随着产品开发方法的成熟，它所能带来的益处将会减到最小。如果筛选过程中的失效数量很小，那么制造过程就很可能在容限范围之内，所发现的故障也可能超出了设计和生产过程的范围。

9.10　闭环根源监测

产品的可靠性需要用一个闭环过程来保证。在产品生命周期的早期阶段，它能为设计和制造提供反馈。从制造、装配、存储、运输、定期维修和使用中获得的数据，健康监测方法有助于对未来设计进行规划和试验；也有助于对产品进行定期维修以保持状态，并阻止灾难性失效。图9.4描绘了用于在整个生命周期对产品进行可靠性管理的闭环式过程。

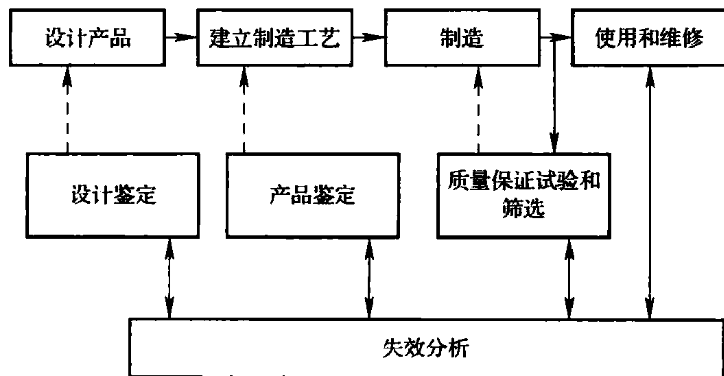


图9.4　在整个生命周期对产品进行可靠性管理的闭环式过程

闭环监测的目标是在产品生命周期中分析所有失效，进而确定失效的根源。根源是最根本的原因或因素，如果失效根源得到修正或清除，将能阻止失效的再次发生。决策失效根源的目的是在根源上确定问题所在，这样就能阻止失效再次发生。在其他产品中，这样做能以最小的代价来确定失效的特征。

在设计、制造和使用过程中，要正确地分析已确定的失效根源，这样就能导致减少现场返修；能节省主要开发成本，提升客户友好度。要记录从每个失效分析中获取到的经验教训，并采取适当的行动，以更新设计、制造工艺和维修行为。

在产品开发完成后，需要把相关资源应用到供应链管理、废弃评估、制造和装配反馈、制造质量保证管理和现场失效的根源分析中。与产品相关的风险分为两类：

① 管理风险：产品开发小组创建管理计划，并对产品的现场性能、制造人员和可制造性实施上述的监测制度。管理风险指这些积极的管理行为所带来的风险。

② 未管理风险：产品开发小组没有进行积极管理而产生的风险。

如果风险管理是必需的，就需要为此制订计划。计划要包括一些细节内容，这些内容与如何监测（收集数据）产品、如何把监测结果反馈到各种产品开发过程中有关。必须要考虑管理过程的可行性、要付出的努力和代价。

9.11 总结

可靠产品的开发不是偶然事件，而是整个产品生命周期中，一个有意识的、系统的、付出辛劳努力后的结果。只有通过健壮的产品设计、使用容限内具有一定能力的工艺过程、使用来自于供应商的质量可靠的零件（材料），才能实现产品的可靠性目标，供应商在容限之内的工艺也要有一定的能力。对所有相关失效机理进行量化理解和建模，可以指导设计、制造过程，并能为试验规范制订计划。

在产品开发早期的概念阶段使用可靠性分析，它有助于决策方案的可行性和风险。在设计阶段，可靠性分析包括零件（材料）选择、设计权衡、公差设计、制造工艺和公差、装配技术、运输和装卸方法、维修和维修性指南的制订等内容。在这些设计分析中，类似于强度、疲劳、断裂、蠕变、公差、腐蚀和老化的工程概念起着重要的作用。把失效物理概念和机械、概率技术结合起来使用，就可以评估潜在问题和权衡，然后根据此内容采取适当的修复行为。

参考文献

IEEE Reliability Society. 1998. IEEE Std. 1332-1998. IEEE standard reliability program for the development and production of electronic systems and equipment.

Mishra, S., and M. Pecht. 2002. In-situ sensors for product reliability monitoring. Proceedings of SPIE 4755: 10-19.

Mishra, S., M. Pecht, T. Smith, I. McNee, and R. Harris. 2002. Remaining life prediction of electronic products using life consumption monitoring approach. Proceedings of the European Microelectronics Packaging and Interconnection Symposium, Cracow, Poland, 136-142, 16-18 June 2002.

Pecht, M., M. Dube, M. Natishan, and I. Knowles. 2001. An evaluation of built-in test. IEEE Transactions on Aerospace and Electronic Systems 37 (1): 266-272.

Ramakrishnan, A., and M. Pecht. 2003. A life consumption monitoring methodology for elec-tronic systems. IEEE Transactions on Components and Packaging Technologies 26 (3): 625-634.

Vichare, N., P. Rodgers, V. Eveloy, and M. Pecht. 2004. In situ temperature measurement of a notebook computer—A case study in health and usage monitoring of electronics. IEEE Transactions on Device and Materials Reliability 4 (4): 658-663, 2004.

练习

9.1 使设计发生改变的零件的生产批量、供应商来源以及零部件特征的变异性都

能使产品的现场使用寿命发生变化，这些因素如何影响那些对可靠性有所影响的设计决策？

9.2 讨论工艺过程控制和应力裕度的相互关系。此关系是如何影响鉴定的？对产品可靠性又有何影响？

9.3 为计算机键盘列出5项典型的生命周期载荷，并描述产品设计如何处理这些载荷，以保证产品的可靠性。

9.4 对于军事应用的关键产品，解释供应链的全球化是如何影响零件选择、管理过程的？

9.5 阐述FMEA和FMMEA的区别，以及这些区别对于可靠性设计的意义。例如一个FMMEA过程是如何影响产品鉴定试验的？

9.6 是否需要把冗余加入到设计中，产品的预期应用如何影响此决策过程？在回答的过程中讨论与产品定义相关的约束条件。

9.7 解释可制造性的概念，此概念如何用于改进产品可靠性？给出一个具体案例。

9.8 与加速试验相比，虚拟鉴定的优点和缺点如何？如何将这些优缺点结合到鉴定项目中，以减少总体设计周期时间？

第 10 章　系统可靠性建模

10.1　引言

本章介绍如何基于产品的零件和子系统对产品可靠性进行建模；可靠性框图，它是一种表达逻辑系统结构的方法，可用来建立系统可靠性模型；以及故障树（Fault Tree）。

10.2　可靠性框图

可靠性框图用以描述系统各单元之间的逻辑关系。第 10.3 节介绍了串联系统，第 10.4 节介绍了冗余系统［包括备用系统、表决（n 中取 k）系统和复杂系统］。本章还利用概率论原理对这些系统结构进行了分析。

10.3　串联系统

在串联系统（Series System）中，所有的子系统都要正常运行，系统才能实现其功能。这意味着任何子系统的失效都将引起系统失效。从可靠性的观点来看，如果所有单元都可靠，那么系统就是可靠的。所谓串联系统，并不一定要求所有的单元都按照串联的方式进行物理连接。串联系统的可靠性框图如图 10.1 所示。每个单元的可靠度用 $R_i(t)$ 表示，其失效时间用 TTF（i）表示。

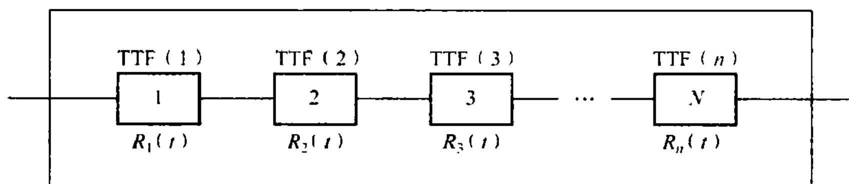


图 10.1　串联系统的可靠度

我们可以根据基本的概率论原理计算系统的可靠度。公式（10.1）是系统可靠度和失效时间的计算方法。在串联系统中，可靠度会随着单元的增多而下降（见图 10.2）。

系统的 TTF：min（TTF（i））（$i=1, 2, \cdots, n$）

$$R_S(t) = R_1(t)R_2(t)\cdots R_n(t) = \prod_{i=1}^{n} R_i(t) \tag{10.1}$$

假定系统每个单元的失效时间都服从指数分布，且失效率恒为 λ_i。那么，每个单

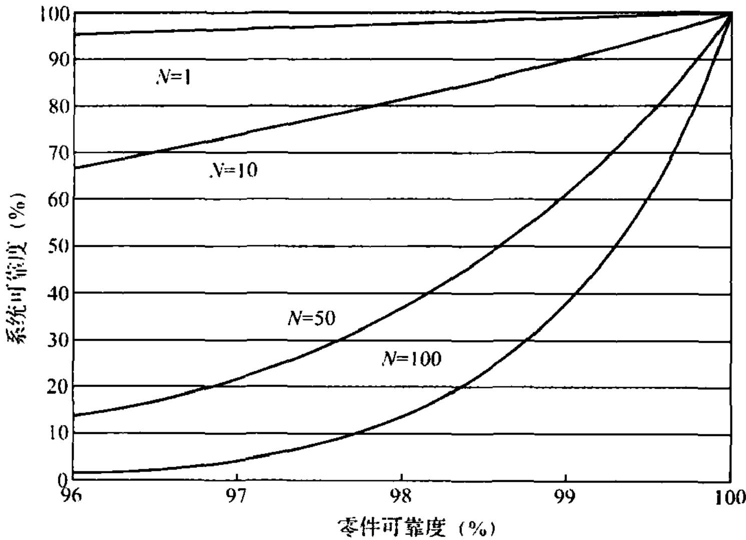


图 10.2　零件的可靠度和数量对串联系统可靠度的影响

元的可靠度为

$$R_i(t) = e^{-\lambda_i t} \tag{10.2}$$

系统的可靠度为

$$R_S(t) = \prod_{i=1}^{n} R_i(t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-(\sum_{i=1}^{n}\lambda_i)t} \tag{10.3}$$

系统的恒定失效率为

$$\lambda_S = \sum_{i=1}^{n} \lambda_i \tag{10.4}$$

系统的平均失效间隔时间（MTBF）为

$$\mathrm{MTBF} = \frac{1}{\lambda_S} = \frac{1}{\sum_{i=1}^{n} \lambda_i} \tag{10.5}$$

如果系统的所有单元都是串联的，且都有恒定的故障率，那么系统的故障率也恒定。恒定故障率和串联系统的假设可以让系统可靠性的数学计算变得更简单，但实际情况通常不是这样的。

案例　10.1

某电子系统包括两个串联零件。假设第 i 个零件的失效由其恒定失效率 λ_i 决定，计算：

① 系统的失效率。

② 系统运行 1000 小时的可靠度。

③ 系统的平均失效时间（MTTF）。

零件的失效率分别为

$$\lambda_1 = 6.5\ 次/10^6\mathrm{h}$$

$$\lambda_2 = 26.0\ 次/10^6\mathrm{h}$$

解：对于恒定失效率，第 i 个零件的可靠度 R_i 为

$$R_i = e^{-\int_0^t \lambda_i(t)dt} = e^{-\lambda_i t}$$

串联系统的可靠度 R_S 为

$$R_S = e^{-\sum_{i=1}^{n}\lambda_i(t)dt} = e^{-\lambda_s t}$$

$$\lambda_S = \sum_{i=1}^{n}\lambda_i$$

对于由恒定失效率的零件组成的串联系统，将给定值代入以上公式，可得：

$$\lambda_S = 32.5 \text{ 次}/10^6\text{h}$$

系统运行 1000h 的可靠度：

$$R_S(1000) = e^{-(32.5\times10^6)\times1000} = 0.968$$

系统的平均失效时间（MTTF）为

$$\text{MTTF} = \int_0^\infty R_{SS}(t)\,dt = \int_0^\infty e^{\lambda_{SS}}\,dt = 1/\lambda_{SS} = 30770\text{h}$$

10.4 冗余系统

当一个或多个零件发生失效，而系统中剩余的零件仍然能够维持其功能的系统称为冗余系统（Redundancy System）。两种常见的冗余是工作冗余（Active Redundancy）和备用冗余（Standby Redundancy）。在工作冗余中，系统运行时，所有零件都处于运行状态。零件的寿命与系统单个单元的寿命消耗速率相同。

而在备用冗余中，冗余零件在系统运行时则不参与其中，只在某些工作零件失效时才起动。与工作冗余相比，备用冗余的零件拥有较长的寿命。

备用冗余包括三种类型：冷备用（Cold Standby）、温备用（Warm Standby）和热备用（Hot Standby）。在冷备用中，直到需要运行前辅助零件都处于关闭状态。这缩短了零件工作的时间，减少了有效寿命的消耗；但在开关零件时，其承受的瞬时应力将会很高，这会加快零件的寿命消耗。在温备用中，辅助零件通常处于起动状态，但却是空闲或空载的。在热备用中，辅助零件组成一个并联的工作系统。热备用零件的寿命消耗与工作零件的寿命消耗速率相同。在热备用系统中，当主系统中的某个零件失效时才需要将备用零件切换到工作环路；这是热备用和工作冗余的不同之处。

10.4.1 工作冗余

工作冗余系统是一个标准的“并联”系统，只有所有单元失效后，它才会失效。某些时候也把并联系统称为 n 中取 1 或（1，n）系统，它表示只要 n 个子系统中的某 1 个能够运行，系统就可以运行。并联系统的可靠性框图如图 10.3 所示。

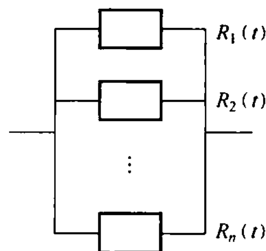


图 10.3 工作冗余系统

在并联系统中，所有单元并不一定都是按照并联的方式进行物理连接的。只有所有的子系统或单元都在 t 时刻失效，系统

才会失效；如果至少有一个单元存活到了系统运行时刻 t，那么系统也就能存活到 t 时刻。并联系统的可靠度可表示为

$$R_S(t)=1-Q_S(t) \tag{10.6}$$

其中，$Q_S(t)$ 是系统失效的概率，它可以表示为

$$Q_S(t)=[1-R_2(t)][1-R_2(t)]\cdots[1-R_n(t)]=\prod_{i=1}^{n}[1-R_i(t)] \tag{10.7}$$

任务时间 t 时刻的系统可靠度为

$$R_S(t)=1-\prod_{i=1}^{n}[1-R_i(t)] \tag{10.8}$$

系统的 TTF：max（TTF（i））（$i=1$，2，…，n）

单元的可靠度和数量对工作冗余系统可靠度的影响如图 10.4 所示。

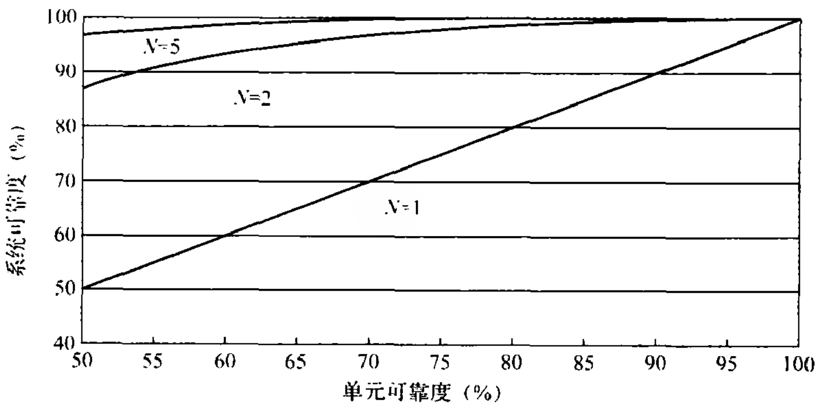


图 10.4　单元的可靠度和数量对工作冗余系统可靠度的影响

通常情况下，每个单元都有不同的失效分布。系统的故障率为

$$\lambda_S(t)=\frac{f_S(t)}{R_S(t)} \tag{10.9}$$

其中，$f_S(t)$ 是系统失效时间概率密度函数（pdf）。系统的平均寿命 m 为

$$m=\int_0^{\infty}R_S(t)=\int_0^{\infty}[1-\prod_{i=1}^{n}(1-R_i(t))]\mathrm{d}t \tag{10.10}$$

例如系统包括两个单元（$n=2$），这两个单元的失效服从指数分布，其失效率恒定不变，分别为 λ_1 和 λ_2，那么系统的平均寿命可通过公式（10.11）计算得到。注意：虽然每个单元的失效率是恒定的，但系统的平均寿命并不等于所有单元失效率之和的倒数，其故障率也会随着时间的推移而变得不恒定。

$$m=\frac{1}{\lambda_1}+\frac{1}{\lambda_2}-\frac{1}{\lambda_1+\lambda_2} \tag{10.11}$$

案例　10.2

如果某电子系统包括两个零件，其失效率恒定，分别为 $\lambda_1=6.5$ 次/10^6h 和 $\lambda_2=26.0$ 次/10^6h。第 i 个零件的失效由其恒定失效率 λ_i 决定。计算：

① 系统运行 1000h 的可靠度。

② 系统的平均失效时间（MTTF）。

③ 系统的失效概率密度函数。

④ 系统的失效率。

解：由于失效率恒定，第 i 个零件的可靠度 R_i 为

$$R_i = e^{-\int_0^t \lambda_i(t)dt} = e^{-\lambda_i t}$$

对于并联系统，则

$$R_P = 1 - \prod_{i=1}^{2}(1 - e^{-\lambda_i(t)}) = e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}$$

失效概率密度函数为

$$f_P(t) = \frac{d[R_P(t)]}{dt} = \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)t}$$

并联系统的失效率为

$$\lambda_P(t) = \frac{f_P(t)}{R_P(t)}$$

将给定值代入以上公式，可得：

$$R_P(1000) = 0.99352 + 0.97434 - 0.96802 = 0.99983$$

并联系统的平均失效时间（MTTF）为

$$MTTF_P = \int_0^\infty R_P(t)dt = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{(\lambda_1 + \lambda_2)} = 161540h$$

失效概率密度函数为

$$f_P(t) = -\frac{d[R_P(t)]}{dt}$$

$$= 6.5\times10^{-6} + 26.0\times10^{-6}e^{-26.0\times10^{-6}t} - 31.5\times10^{-6}e^{-31.5\times10^{-6}t}$$

将所得值代入前面的公式，可以计算并联系统的失效率。

10.4.2　备用系统

备用系统包括一个处于工作状态的单元或子系统以及一个或多个未工作单元。工作单元失效时，将起动这些未工作单元。如果工作单元失效，故障传感器将发出一个信号，并通过转换开关打开备用单元。最简单的备用系统如图 10.5所示，它由两个单元组成。一般情况下，系统拥有 n 个单元时，则有（$n-1$）个单元处于备用状态。

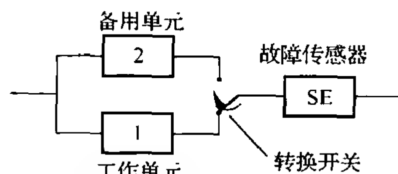


图 10.5　备用系统

当工作单元和备用单元拥有相等的恒定失效率 λ，系统有表现良好的开关和传感器，即 $\lambda_{sw}=0$，那么此类系统的可靠度函数为

$$R(t) = e^{-\lambda t}(1 + \lambda t) \tag{10.12}$$

10.4.3　表决系统

只要 n 个单元中有 k 个以上单元正常时，系统就正常工作，这样的系统称为 n 中取 k

或（k，n）系统［也叫表决系统，（k，n）系统］。n 中取 k 系统的可靠性框图与并联系统的相同。但对于 n 中取 k 系统，至少要有 k 个单元正常运行，系统才能实现其功能（见图10.6）。

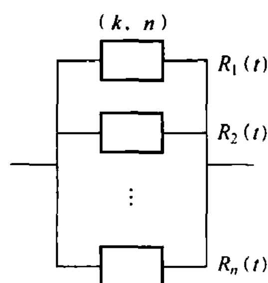


图10.6　n 中取 k 系统

当系统的单元拥有不同的失效分布时，系统可靠度函数就很难用数学计算公式来表达。假设所有单元的失效分布都相同，即均为 $Q(t)$，系统的可靠度可以通过二项式分布来计算：

$$R_S(t) = \sum_{i=k}^{n}\binom{n}{i}[1-Q(t)][Q(t)]^{n-1} \tag{10.13}$$

表决系统的失效概率为

$$Q(t) = 1 - R(t) = 1 - \sum_{i=k}^{n}\binom{n}{i}[1-Q(t)][Q(t)]^{n-i} = \sum_{i=0}^{k-1}\binom{n}{i}[1-Q(t)][Q(t)]^{n-i} \tag{10.14}$$

概率密度函数为

$$Q(t) = \frac{\mathrm{d}Q_S(t)}{\mathrm{d}t} = \frac{n!}{(n-k)!\ (k-1)!}[1-Q(t)]^{k-1}[Q(t)]^{n-k}Q(t) \tag{10.15}$$

故障率可计算为

$$\lambda_S(t) = \frac{f_S(t)}{R_S(t)} \tag{10.16}$$

案例　10.3

某工作冗余系统为3中取2系统，所有单元具有相同的可靠度 R，计算系统的可靠度。

解：此处，$n=3$，$k=2$。从公式（10.13）知，n 中取 k 冗余系统的可靠度为

$$R_{2out\ of 3} = \frac{3!}{(1!)(2!)}R^2Q^1 + \frac{3!}{(0!)(3!)}R^3Q^0$$
$$= 3R^2(1-R) + R^3$$

两个单元正常，一个单元失效的概率　　三个单元都正常工作的概率

10.4.4　冗余的限制因素

一些操作和设计问题往往会使冗余难以发挥其优势。其中的三个问题分别是共模失效、载荷分担失效以及开关和备用单元失效。

两个或更多冗余零件相互关联的现象会引发共模失效（Common Mode Failure），这种现象将会使这些冗余零件同时发生失效。有很多原因会导致共模失效（例如共电连接、共享的环境应力以及共同的维修问题）。在系统可靠性分析中，共模失效产生的影响与在并联系统中加入额外的零件而产生的后果是相同的。

载荷分担失效是由于某个零件的失效增加了其他零件的应力级别而产生的。应力级

别的增加会影响工作零件的寿命。对于冗余的发动机、电动机、泵、机构和许多其他系统，以及在并联系统中工作的设备来说，一个零件的失效可能会增加其他零件的载荷，进而会缩短系统的失效时间（或增加它们的故障率）。

通常，对备用系统中的开关和传感器进行一些假设。对于开关单元，假定其为单向，只有在控制器发出转换指令时才会有所响应，开关的失效只会发生在工作状态下。对于备用单元，通常的假设未运行的备用单元在不通电时将不会失效。当所有这些假想条件中的任意一个不能满足时，开关和备用单元将发生失效。监视器或传感器的失效包括动态（在需要转换时，开关不响应）失效和静态（在不需要转换时的开关动作）失效。

10.4.5　复杂系统

如果某个系统在结构上不能分解成串、并联结合的形式，那么我们认为它属于复杂系统。有三种方法可以对复杂系统进行可靠性分析。下面将以图 10.7 所示系统为例，介绍这三种方法。

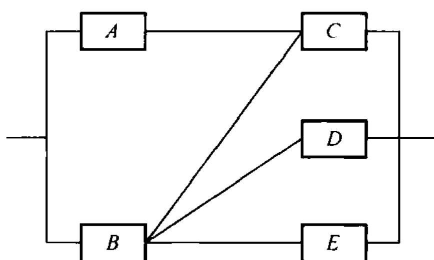


图 10.7　复杂系统

完全枚举法（Complete Enumeration Method）的基础是列出失效单元所有可能的组合形式。图 10.7 所示系统所有可能的状态都列在了表 10.1 中。其中，○表示系统处于正常运行状态，F 表示系统处于失效状态；大写字母表示某个单元处于运行状态，小写字母表示某个单元处于失效状态。

表 10.1　完全枚举法示例

系统描述	系统条件	系统状态
所有单元都可运行	*ABCDE*	○
一个单元处于失效状态	*aBCDE*	○
	AbCDE	○
	AbcDE	○
	ABCdE	○
	ABCDe	○
两个单元处于失效状态	*abCDE*	F
	aBcDE	○
	aBCdE	○
	aBCDe	○
	AbcDE	F
	AbCdE	○
	AbCDe	○
	ABcdE	○
	ABcDe	○
	ABCde	○
三个单元处于失效状态	*ABcde*	F
	AbCde	○
	AbcDe	F
	AbcdE	F
	aBCde	○
	aBcDe	○
	abCDe	F
	abCdE	F
	abcDE	F
四个单元处于失效状态	*Abcde*	F
	aBcde	F
	abCde	F
	abcDe	F
	abcdE	F
五个单元都处于失效状态	*abcde*	F

每一种代表系统状态的组合都可以用该单元处于某特定状态概率的乘积来表示。例如组合形式2可以写成 $(1-R_A)\ R_BR_CR_DR_E$，其中，$(1-R_A)$ 表示单元 A 在 t 时刻失效的概率。对于处于运行状态○的系统，其可靠度可以表示为所有这些组合之和即

$$
\begin{aligned}
R_S = {} & R_AR_BR_CR_DR_E + (1-R_A)R_BR_CR_DR_E + R_A(1-R_B)R_CR_DR_E + \\
& R_AR_B(1-R_C)R_DR_E + R_AR_BR_C(1-R_D)R_E + R_AR_BR_CR_D(1-R_E) + \\
& (1-R_A)R_B(1-R_C)R_DR_E + (1-R_A)R_BR_C(1-R_D)R_E + \\
& (1-R_A)R_BR_CR_D(1-R_E) + \\
& \cdots \qquad + \\
& (1-R_A)R_B(1-R_C)(1-R_D)R_E
\end{aligned} \tag{10.17}
$$

简化后，系统可靠度可表示为

$$
\begin{aligned}
R_S = {} & R_BR_CR_DR_E - R_AR_BR_C - R_BR_CR_D - R_BR_CR_E - \\
& R_BR_DR_E + R_AR_C + R_BR_C + R_BR_D + R_BR_E
\end{aligned} \tag{10.18}
$$

条件概率法（Conditional Probability Method）的基础是全概率法则（Law of Total Probability），它允许在 t 时刻把系统分解为特定的单元和状态。例如系统在 t 时刻的可靠度等于单元 A 处于运行状态时的系统可靠度（表示为 $R_S \mid A_G$）乘以单元 A 的可靠度，再加上单元 A 处于失效状态时的系统可靠度（$R_S \mid A_B$）乘以单元 A 的不可靠度，为

$$R_S(R_S|A_G)R_A + (R_S|A_B)Q_A \tag{10.19}$$

将这种分解过程一直延续到所有单元，列出每个单元的可靠度和不可靠度。以图10.8所示系统为例，用单元可靠度表示的系统可靠度为

$$R_S(R_S|C_G)R_C + (R_S|C_B)Q_C \tag{10.20}$$

如果单元 C 在 t 时刻运行，我们就可以对系统配置进行简化，如图10.8所示。因此，系统的可靠度就等于前面给出的串、并联结合系统的可靠度，或

$$R_S|C_G = [1-(1-R_A)(1-R_B)] \tag{10.21}$$

如果单元 C 在 t 时刻失效，系统也可以进行简化，如图10.9所示。此时，系统的可靠度为

$$R_S|C_B = R_B[1-(1-R_D)(1-R_E)] \tag{10.22}$$

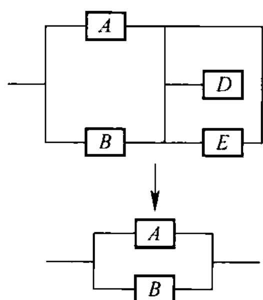


图10.8　单元 C 运行时系统的简化形式

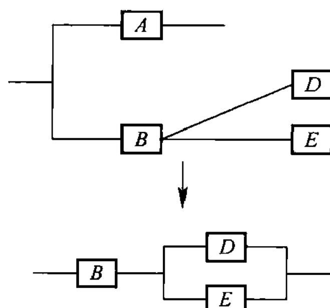


图10.9　单元 C 失效时系统的简化形式

把公式（10.21）和公式（10.22）代入公式（10.19）中，就可以得到系统的可靠

度为

$$R_S = (R_S|R_G)R_C + (R_S|R_B)Q_C$$
$$= [1-(1-R_A)(1-R_B)]R_C + R_B[1-(1-R_D)(1-R_E)](1-R_C) \quad (10.23)$$

系统可靠度可以由其单元的可靠度来表示。公式（10.23）的简化形式与公式（10.18）相同。用前面介绍的方法可以获得单元（也就是每个框图）的可靠度。

割集（Cut Set）是由系统中部分单元组成的集合。当集合中的全部单元都失效时，系统就发生失效。能够导致系统失效的最少单元集合称之为最小割集（Minimal Cut Set）。如果从最小割集中把某个单元移除（也就是将它考虑成不失效），系统将不会失效。这意味着所有来自最小割集中的单元都必须失效才能引起系统失效。利用最小割集计算系统可靠性的步骤如下：

① 为给定系统确定最小割集。

② 将割集中的每个单元按并联结构进行建模。

③ 用串联结构将所有最小割集进行连接。

④ 将系统建模为若干串联形式的割集，每个割集中的单元成并联形式，以此求解系统的可靠度。

对于前面提到的示例，可以确定以下割集：

$$C_1 = \{A, B\}$$
$$C_2 = \{B, C\} \quad (10.24)$$
$$C_3 = \{C, D, E\}$$

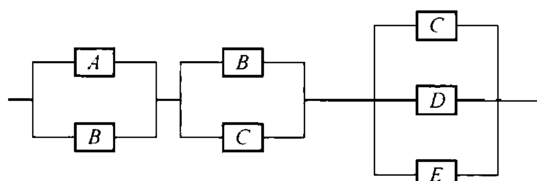


图 10.10　用最小割集表示的系统框图

按照上面描述的步骤就可以用最小割集来表示系统框图，如图 10.10 所示。

利用串联和并联系统的计算方法，系统的可靠性可以表示为

$$R_S = [(1-R_A)(1-R_B)][1-(1-R_B)(1-R_C)] \\ [1-(1-R_C)(1-R_D)(1-R_E)] \quad (10.25)$$

经过简化，可以得到以下结果，它与公式（10.18）相同。

$$R_S = R_AR_C + R_BR_C + R_BR_D + R_BR_E - R_AR_BR_C - R_BR_DR_E - \\ R_BR_CR_D - R_BR_CR_E + R_BR_CR_DR_E \quad (10.26)$$

10.5　故障树分析

故障树分析（Fault-Tree Analysis，FTA）是一种用来确定失效的潜在原因和估计失效概率的演绎方法。故障树描述了产品的设计失效和潜在失效，通过演绎的方法追踪系统失效。它图形化地表示了系统的功能和行为，每一次分析都是针对某一个系统失效进行定性和定量的可靠性分析。故障树的目的就是为了展现导致顶事件（Top Event）发生的事件集合，特别是那些主要的失效事件。

失效可以通过很多方式来划分（硬件故障或人为差错。硬件故障包括：早期故障、

随机故障或老化故障，主要故障、次要故障或命令故障，主动故障或被动故障）。在本节中，我们不区别由这些不同分类方法而造成的差异［Lewis，1996］。

案例　10.4

图10.11是一个断电系统的可靠性框图。如果外部电源和应急电源都失效，将发生断电。如果电压监控器或柴油发电机失效，紧急电源将失效（当外部电压低于下限值时，电压监控器发送信号起动柴油发电机）。

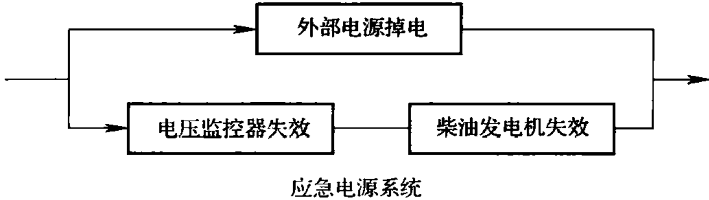


图10.11　某断电系统的可靠性框图

断电系统的故障树如图10.12所示。

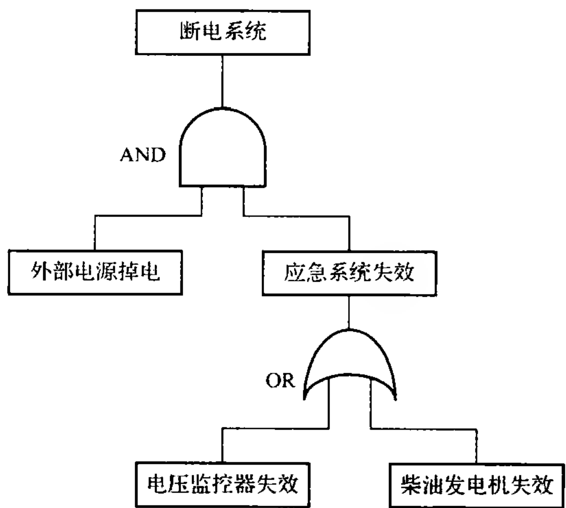


图10.12　断电系统的故障树

10.6　故障树分析的步骤

故障树分析包括三个步骤：

① 绘制逻辑框图或使用故障树元素建立故障树。此阶段要求对系统进行完全定义，并了解其运行过程。分析每个失效的可能引发原因和影响，并将其与顶事件联系起来。

② 将逻辑代数应用于逻辑框图，并找出事件间的代数关系。尽可能用逻辑代数形式来简化表达此关系。

③ 用概率方法计算每个中间事件和顶事件发生的概率。必须要获得每个事件的发

生概率，也就是需要考虑每个单元或者子系统在每种失效模式下的可靠性。

这些用来构建故障树的图形符号分为两种类型：逻辑门符号和事件符号。基本的逻辑门符号包括与门、或门、n 中取 k 的表决门、顺序（优先）与门、异或门和禁门等。基本的事件符号包括基本事件、未展开事件、条件时间、触发事件、结果事件和转入、转出事件等［Lewis，1996；Rao，1992；Kececioglu，1991］。故障树的量化分析包括根据故障树各事件间相互影响的逻辑表达计算顶事件发生的概率。建立故障树的常用符号如表 10.2 所示，量化分析的逻辑表达如表 10.3 所示。

表 10.2　故障树符号——事件和逻辑门

符号	名称	含义
	方形	故障事件
	圆形	独立的主要故障
	菱形	未展开事件或未探明事件
	房型	预期正常发生的事件（触发事件或开关事件）
	与门	所有输入为真时，输出为真
	或门	任何输入为真时，输出都为真
X A	禁门	满足条件时，存在输出
A A	三角转入/转出	事件转移——跳转到树的另一部分

表 10.3　与和或关系的逻辑代数表达

A	B	A AND B	A AND B	A OR B	A NOR B
0	0	0	1	0	1
1	0	0	1	1	0
0	1	0	1	1	0
1	1	1	0	1	0

案例 10.5

分析以下故障树（见图10.13）。

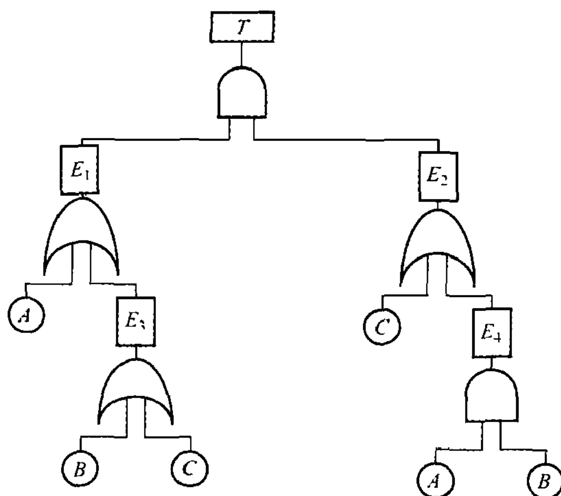


图10.13 某故障树

自顶向下计算：

① $T=E_1\cap E_2$

② $E_1=A\cup E_3$；$E_2=C\cup E_4$。

③ $E_3=B\cup C$；$E_4=A\cap B$。

④ $T=(A\cup E_3)\cap(C\cup E_4)=[A\cup(B\cup C)]\cap[C\cup(A\cap B)]$。

由底向上计算：

① $E_3=B\cup C$；$E_4=A\cap B$

② $E_1=A\cup E_3$；$E_2=C\cup E_4$

③ $E_1=A\cup(B\cup C)$

④ $E_2=C\cup(A\cap B)$

⑤ $T=E_1\cap E_2=[A\cup(B\cup C)]\cap[C\cup(A\cap B)]$

任何一种计算方法都可用于故障树分析：

结合律：$A\cup(B\cup C)=(A\cup B)\cup C$

交换律：$(A\cup B)\cup C=C\cup(A\cup B)$

因此，$T=[C\cup(A\cup B)]\cap[C\cup(A\cap B)]$

分配律：$T=C\cup[(A\cup B)\cap(A\cap B)]$

$A\cap B=B\cap A$

结合律：$T=C\cup[(A\cup B)\cap B\cap A]$

吸收律：$(A\cup B)\cap B=B$

因此，故障树可以简化为图10.14所示的形式，仅当C发生或当A和B都发生时，T发生。

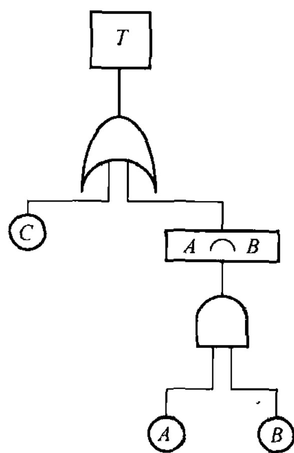


图10.14 简化后的故障树

参考文献

Kececioglu, B. D. 1991. Reliability engineering handbook, Vols. 1-2. Englewood Cliffs, NJ: Prentice Hall.

Lewis, E. E. 1996. Introduction to reliability engineering. New York: John Wiley & Sons.

Rao, S. S. 1992. Reliability-based design, 505-543. New York: McGraw-Hill.

练习

10.1 考查图 10.15 所示的两个可靠性框图。它们拥有相同数量相同内容的单元。假定每个单元的可靠度都为 RX，其中，X 为单元的名称。

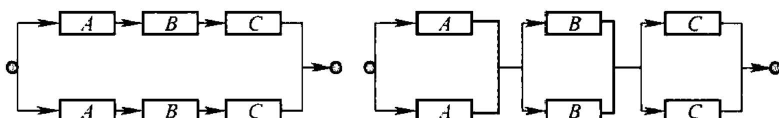


图 10.15 10.1 题的两个可靠性框图

(a) 当所有单元的失效都相互独立时，这两种结构的系统可靠性是否存在差异？对给出的答案进行解释。不必给出完整的数学推导过程。

(b) 哪种结构对共模失效更为敏感？为什么？假定每个单元（A、B 和 C）的失效机理都不同，这些机理受到不同载荷的影响。

10.2 以下图 10.16 所示的可靠性框图属于复杂系统，它不能被分解为“串——并联”结构。请使用条件概率方法推导该系统的可靠性方程。使用 B 作为分解单元。绘制“B 运行”和“B 失效”这两种条件下的可靠性框图。

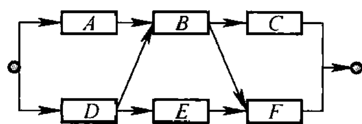


图 10.16 10.2 题的可靠性框图

10.3 考查图 10.17 所示的可靠性框图，并推导出它的可靠性方程。RX 表示系统每个单元的可靠度，其中 X 为单元的名称。第 3 部分（4 个并联的单元 C）是一个 4 中取 2 系统——也就是只要有 2 个单元运行，系统就可以运行。

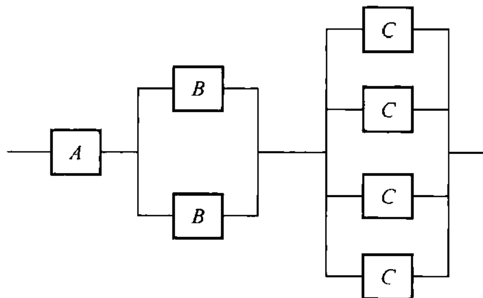


图 10.17 10.3 题的可靠性框图

10.4 请推导出图10.18所示系统的可靠性方程（手动推导）。注意：它是一个复杂系统。

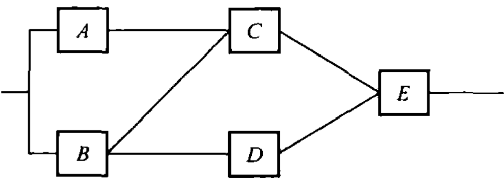


图10.18 10.4题的可靠性框图

求解：

(a) 系统在100h处的可靠度。

(b) 系统在0h处的可靠度。

(c) 1000h处的失效率。

(d) 系统开始出现损耗的时间区域（用图形表示）。

(e) 75%的系统出现失效的时间是多少？

单　元	失效分布	参数（小时或等式）
A	Weibull 三参数分布	$\beta=3$，$\eta=1000$，$\gamma=100$
B	指数分布	平均实效时间间隔 +1000
C	对数分布	中值 =6，标准偏差 =0.5
D	Weibull 三参数分布	$\beta=0.7$，$\eta=150$，$\gamma=-100$
E	正态分布	中值 =250，标准偏差 =15

(f) 如果把 *C* 和 *D* 进行换位，将会产生什么样的结果？

10.5 对于顶事件 *T*，其最小割集为：*ABC*、*BDC*、*AE*、*ADF* 和 *BEF*。绘制出这些最小割集的顶事件的故障树。

第11章 冗余和容错产品的可靠性分析

尽管产品的一些部件发生失效，但是容错产品仍然可以正常运行。这种容错经常通过使用冗余来实现——也就是通过提供替代方式来完成既定任务。本章将介绍评估容错产品可靠性的方法，首先讨论简单冗余产品，这类产品拥有重复部件，具有固定的失效概率；然后介绍一些稍微复杂的冗余技术，例如备用冗余、多数表决和混合冗余，这些冗余的失效概率都与时间相关；还将介绍更复杂的关联失效、多模失效等。本章还将介绍对于嵌入式容错计算机产品非常重要的覆盖建模技术，同时还要介绍用于分析巨大模型的特殊技术和一些高级主题。

1. 术语

部件（Component）：基本的产品单元，部件通常是使用现场可替换或者可修理的最小产品元素。它要么正常运转，要么失效。一个冗余部件至少要有一个重复功能。

产品（Product）：一个整合冗余和非冗余部件，并具有某种功能的集合。产品可靠性分析用于估计产品中正常运行的部件能使产品实现其特定功能的概率。

路径集（Path Set）：冗余的基本单元；一种完成实际任务的方法。假如所有与路径集相关的部件是可运行的，那么产品就是可运行的。当一个部件失效时，一个与之相关的路径也失效，但是冗余路径的失效并不一定导致产品失效。

割集（Cut Set）：一系列部件的失效会导致产品失效。假如所有与割集相关的部件都失效，那么产品就会失效。即使割集中的全部或部分部件是可运行的，产品也不一定是可运行的。

2. 符号

部件通常用 A、B、C 等符号标记。部件正常或失效的概率表述如下：

A——部件 A 运转正常；

$\overline{A}$——部件 A 失效；

P_i——部件 i 可运行的概率，也就是 i 的可靠度；

$P_i(t)$——部件 i 在 t 时刻可运行的概率，也就是 i 在 t 时刻的可靠度；

q_i——部件 i 已失效的概率，也就是 i 的不可靠度；$q_i(t)$ 表示部件 i 在 t 时刻失效的概率，也就是 i 在 t 时刻的不可靠度。

与冗余配置相关概率表述使用以下符号：

R，$R(t)$ ——产品的可靠度（可能与时间相关）；

U，$U(t)$ ——产品的不可靠度（可能与时间相关）[对于所有的 $t \geqslant 0$，$R+U=1$，$R(t)+U(t)=1$]；

A，$A(t)$ ——产品的可用度（可能与时间相关）。

11.1 静态冗余——组合建模

一些产品通过使用静态冗余（常称为被动冗余）来实现容错。在这些产品中，部件的运行和相互关联状态不会因某个失效部件而发生变化。下一节我们讨论动态冗余产品。在动态冗余产品中，某个部件发生失效时，产品将重组。

11.1.1 简单冗余

实现容错最简单的方法是重复，两个冗余部件都用来确保一项功能。图11.1为重复性的单个部件。这个产品包括两个冗余部件，只要其中一个正常，产品就能正常运行。因此，产品的可靠度由两个部件中至少一个正常运行的概率决定——另一种概率则是两个部件都失效。

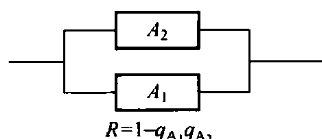


图11.1 重复性单个部件

产品可以拥有多个冗余部件，图11.2是某需要A和B两个部件的产品实现冗余的两种方法。第一种结构提供了两条冗余路径，这两条路径对于产品来说是逻辑对等的(Logically Equivalent)。当每条路径上至少一个部件发生失效时，这种结构也将发生失效。第二种结构分别将每个部件进行复制，提供了四条不相交的冗余路径，只要A_i或B_i运行，这种结构就是有效的（稍后将讨论这些结构的可靠性分析）。

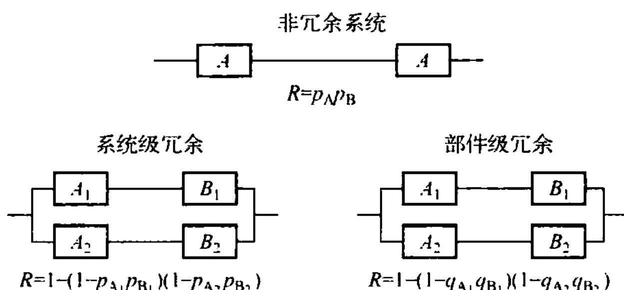


图11.2 简单冗余系统的功能重复

评估简单冗余结构的第一步是绘制可靠性框图(Reliability Block Diagram)，以表示确保产品正常运行的部件组合形式。可靠性框图中的每个节点代表一个部件，部件之间的连线表示部件之间的逻辑关系。从起点到终点间的路径表示部件的一种组合形式。假如一条路径上的所有部件都是可运行的，那么产品就是可运行的。能保证产品正常运行的各种路径形式都可以从可靠性框图中确定。当一条路径上的一个或多个部件发生失效的时候，这条路径就不再可用。因为冗余表示至少要有一条路径能保证产品运行，所以简单冗余产品的可靠度就是至少有一条路径可用的概率。

许多可靠性框图可以分解成部件的串联或并联组成形式。它们的可靠度是相互独立的，将它们的可靠度结合起来，就可以估算出整个产品的可靠度。

1. 串联结构

一个产品由n个部件组成，如果任何一个部件的缺失都让产品无法运行，那么从可靠性的角度来讲，此产品就是一个由n个元素组成的串联产品。这种产品的可靠性框图如图11.3所示。基于部件独立失效的基本假设，串联产品的可靠度就是路径上每个元素都正常运行的概率：



图11.3 由n个部件组成的串联系统

$$R_{\text{series}} = \prod_{i=1}^{n} p_{A_i}(1) \tag{11.1}$$

2. 并联结构

当产品的正常运行仅需要众多部件中的一个正常运行时，就要使用部件的并联结构形式。拥有 n 个元素的并联产品的可靠性框图如图 11.4 所示。因为产品的运行只需要 n 个部件中至少有一个可用即可，所以会产生 n 条不同的可用路径。当 n 个部件全部失效时，产品才会失效。并联产品的可靠度表示为

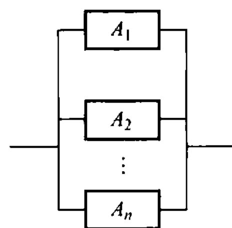


图 11.4 包括 n 个部件的并联系统

$$R_{\text{parallel}} = 1 - \prod_{i=1}^{n} q_{A_i} \tag{11.2}$$

3. 串并联结构

对于由串联和并联部件构成的产品，先分析串联和并联子集，然后将分析结果综合起来，就可得出此类产品的可靠度。每个串联或并联结构都可用可靠性框图中的伪部件（Pseudo Component）进行分析或替代，伪部件的可靠度必须与它所代替的子集拥有相等的可靠度。如图 11.2 所示的两个冗余系统就是简单的串并联结构。

在图 11.5 中，系统级冗余结构被分解成由两个串联连接组成的并联连接。两个串联连接又被伪部件 C_i 代替，C_i 的可靠度等价于 A_i 和 B_i 串联的可靠度。最终，产品就简化成了由 C_1 和 C_2 并联构成的简单系统。再将并联简化，就可以得到如图 11.2 所示的单个系统。

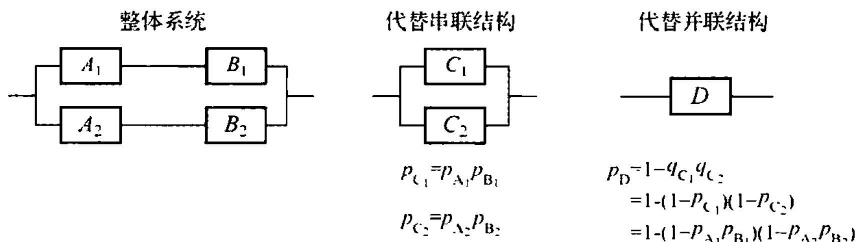


图 11.5 系统级冗余结构的可靠度估计

在图 11.6 中，由图 11.2 而来的部件级冗余结构被分解成由包括两个并联结构的串联结构。并联的部件 A_i 和 B_i 又分别被单个部件 C_i 代替，C_i 的可靠度等同于并联结构的可靠度。串联部件 C_1 和 C_2 又被简化成单个部件 D，D 的可靠度与图 11.2 所示的可靠度相同。

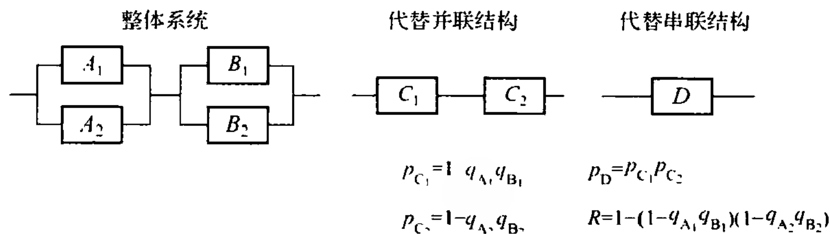


图 11.6 部件级冗余结构的可靠度估计

假定图11.2中的部件可靠度是相等的（$p_{A1}=p_{A2}=p_{B1}=p_{B2}$），就可以对图中部件级和系统级冗余的可靠度进行比较。结果如图11.7所示，对于简单的产品，部件级冗余所能达到的可靠度比系统级冗余中所有部件的可靠度都高。

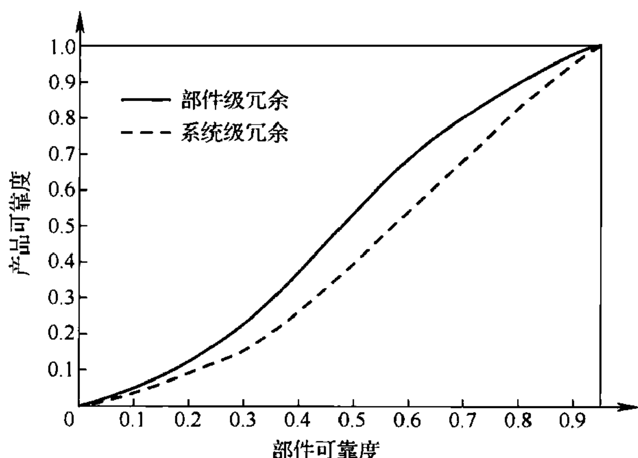


图11.7　两种简单冗余的比较

4. 非串并联结构

一种更广泛实用的分析冗余产品可靠度的方法需要确定产品的路径集，并计算这些路径集的可靠度。这种方法更加普遍适用，因为它更适合那些无法简化为串联或者并联结构的产品。

路径集是这样一种部件集合：如果此集合中的所有部件都可运行，那么产品就可运行。最小路径集是路径上可使产品运行的最小元素数量。假如产品有 n 组最小路径集 p_j（$j=1$，…，n），那么产品的可靠度就是至少一条路径中元素运行的概率：

$$R = Prob\left\{\begin{matrix} n \\ i=1 \end{matrix} p_j\right\} \tag{11.3}$$

系统级冗余如图11.2所示，它的最小路径集是 $\{A_1, B_1\}$ 和 $\{A_1, C_2\}$。其他非最小路径集包括 $\{A_1, B_1, B_2\}$ 和 $\{A_1, A_2, B_1, B_2\}$。公式（11.3）可用来确定系统级冗余的可靠度：

$$\begin{aligned} R &= Prob\{A_1B_1 \cup A_2B_2\} \\ &= Prob\{A_1B_1\} + Prob[A_2B_2] - Prob\{A_1A_2B_1B_2] \\ &= p_{A_1}p_{B_1} + p_{A_2}p_{B_2} - p_{A_1}p_{A_2}p_{B_1}p_{B_2} \\ &= 1-(1-p_{A_1}p_{B_1} - p_{A_2}p_{B_2} + p_{A_1}p_{A_2}p_{B_1}p_{B_2}) \\ &= 1-(1-p_{A_1}p_{B_1})(1-p_{A_2}p_{B_2}) \end{aligned} \tag{11.4}$$

结果与图11.6所示的结果相同。

假设有图11.8所示产品，此产品因为部件 A_3 而不能被简化为一个串并联结构。可根据 A_3 的状态（即运行或失效）将此图简化为两个简单的结构图。即产品的可靠度可由式（11.5）计算：

$$R_{sys}=Pr\{系统运行|A_3\ 运行\}Pr(A_3\ 运行)+Pr\{系统运行|A_3\ 失效\}$$
$$=R_1p_{A_3}+R_2q_{A_3} \tag{11.5}$$

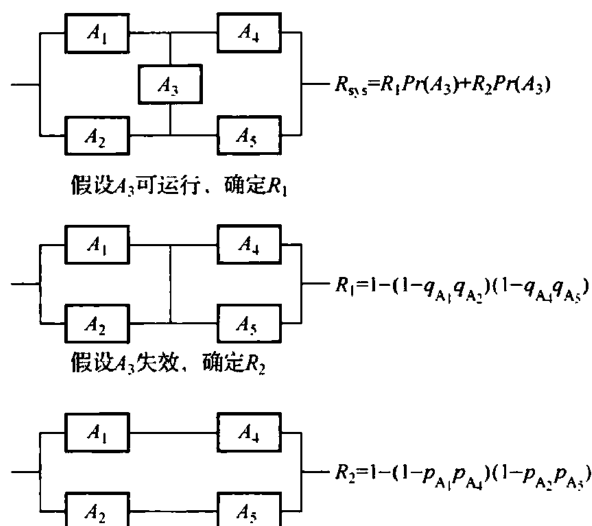


图 11.8 非串并联系统的可靠度估计

假设 A_3 运行，那么图 11.8 中的产品就可以简化为由两个并联结构组成的串联结构（类似于图 11.6），其可靠度为

$$R_1=(1-q_{A_1}q_{A_2})(1-q_{A_4}q_{A_5}) \tag{11.6}$$

同样，假设 A_3 失效，那么图 11.8 中的产品就可以简化为由两个串联结构组成的并联结构（类似于图 11.5），其可靠度为

$$R_2=1-(1-p_{A_1}p_{A_4})(1-p_{A_2}q_{A_5}) \tag{11.7}$$

11.1.2 掩蔽冗余

在拥有重复功能部件的产品中，重复部分可用于检测错误，当两个冗余部件不能产生相同输出时（在有通信连接的情况下），那么其中一个就是有问题的。但是，要想知道哪一个输出是正确的，还需要有额外的信息。假设产品有三组或更多的冗余部件，你该如何确定哪个是有问题的？多数表决器（Majority Voter）可用来判定好的输出和失效单元。它把那些经常出现的输出当做正确输出，同时把那些不同于大多数输出的输出当做来自于失效部件的不正确输出。只要大多数部件正常运行，那么它就可以识别正确的输出。

1. 三重模块冗余

图 11.9 所示为一种遮掩冗余，或称为三重模块冗余（Triple Modular Redundancy，TMR）。在 TMR 产品中，三个冗余部件执行完全相同的任务，表决器从三个冗余的输出中选出正确输出。只要其中两个冗余部件正常运行，并且表决器没有失效，那么这个 TMR 结构就可正常运行。

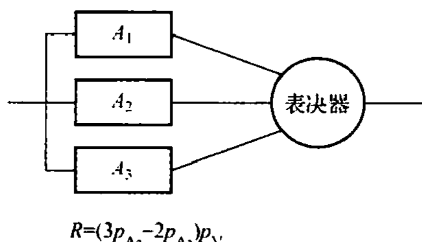


图 11.9 TMR 系统的可靠度

两个单元的运行状态可组成三种情况，设每一种情况发生的概率都是p_A^2，设三个部件都可运行的概率为p_A^3。则TMR结构的可靠度由式（11.8）计算：

$$R_{TMR}=r_V[p_A^3+3p_A^2(1-p_A)]=p_V(3p_A^2-2p_A^3) \tag{11.8}$$

其中，p_V表示表决器正确运行的概率，p_A是每个冗余单元的可靠度。

每个单个部件的可靠度和表决器的可靠度决定了TMR产品是否改进了单个部件的可靠度。对于表决器可靠度的三个不同值，TMR产品的可靠度是单个部件可靠度的函数。图11.10中的对角线表示单个部件的可靠度。

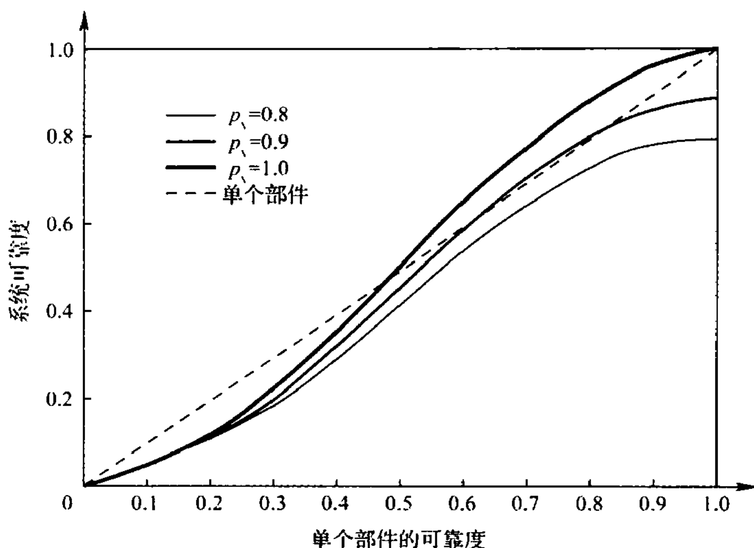


图11.10　三重模块冗余

当表决器表现良好，单个部件的可靠度大于0.5时，TMR产品仅比使用单个部件的可靠度高一些。假如表决器的可靠度仅为0.9，那么TMR产品的可靠度将比单个部件的可靠度低。在大多数实际产品中，表决机制比重复功能部件简单，因而MTR产品失效的概率更低，它是一种改进可靠度的有效方法。

2. N重模块冗余

TMR的概念可以扩展为一种更高级别的冗余。N重模块冗余（N-Modular Redundant，NMR）拥有$N=2n+1$个冗余单元（N是奇数）。如同在TMR产品中一样，NMR中的表决器选择超过半数单元的输出作为正确输出。只要失效的冗余部件数小于n，系统就可以正常运行。NMR产品的可靠度由式（11.9）计算：

$$R_{TMR}=p_V\times\left[\prod_{i=n+1}^{2n+1}\binom{2n+1}{i}p_A^i(1-p_A)^{2n+1-i}\right] \tag{11.9}$$

其中，p_V是表决器的可靠度，p_A是单个部件的可靠度。

图11.11对比了三种不同NMR结构（$N=3$，5，7）的可靠度（假设其中的表决器都没有失效）与单个部件的可靠度。在所有情况下，只要单个部件的可靠度超过0.5，NMR都会使系统的可靠度有所改进。当设定单个部件的可靠度为0.85时，图11.12显

示了 3 种 NMR 产品可靠度与表现不完全良好的表决器的函数关系。当使用单个部件可靠度 $p_A = 0.85$ 的 TMR 产品（$N = 3$）时，表决器的可靠度必须超过 0.91 才能达到改进系统可靠度的目的。

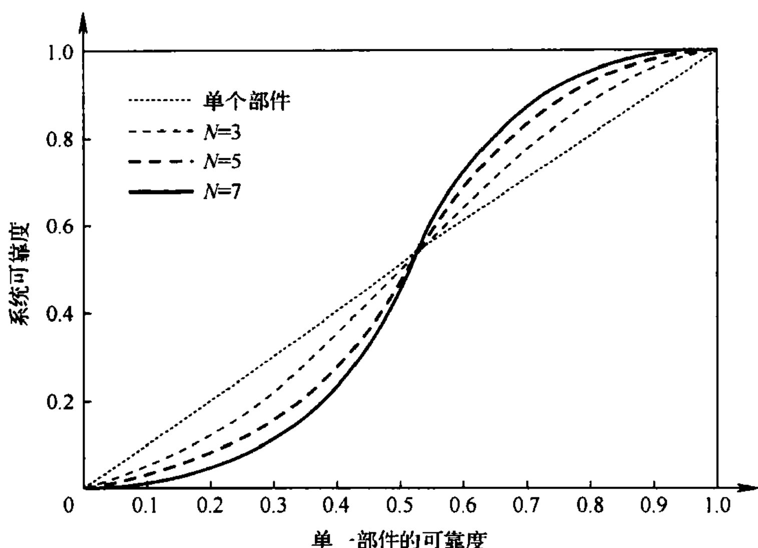


图 11.11 拥有无失效表决器的 NMR 的可靠度

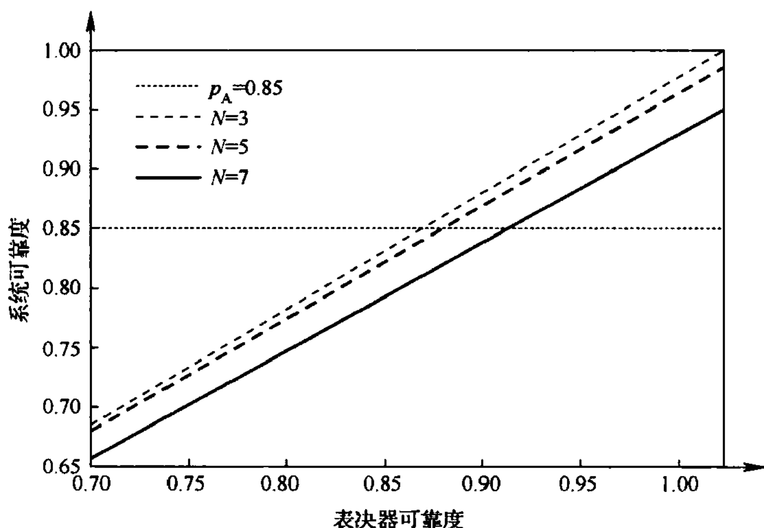


图 11.12 部件可靠度为 $p_A = 0.85$ 时表决器的不可靠度产生的影响

对于五重模块冗余产品，p_V 应该大于 0.88。在七重模块产品中，表决器的可靠度应该大于 0.86，才可以改进单个部件的可靠度。

从可靠性的角度来讲，表决器是一种串联连接，假如它的可靠度没有单个部件的可靠性高，那么 NMR 产品将不会比单个部件更可靠。

11.1.3　故障树

故障树（Fault-Tree）模型是产品失效准则的一种逻辑表示。故障树的顶事件（Top Event）一般描述为所分析产品的失效，并把顶事件分解成引发顶事件的原因事件。这些原因事件用逻辑门（例如与门、或门或 n 中取 m 门）连接起来（当 n 个输入事件中有 m 发生时，则 n 中取 m 门为真）。图11.13所示即是一个故障树的示例，它表示的是来自于图11.2的部件级冗余系统。顶事件（产品失效）由 A_1 和 A_2 的失效引起，或者由 B_1 和 B_2 的失效引起。故障树的底事件（也就是 A_i 和 B_i）是故障树的基本事件，通常用来表示部件失效。

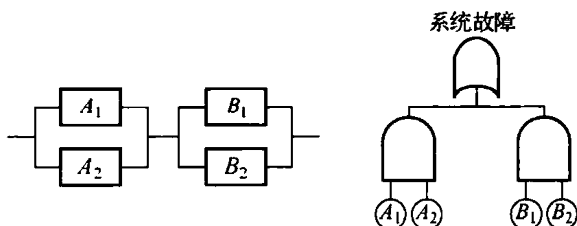


图11.13　等效的可靠度框图和故障树

故障树分析要根据已知的基本事件的发生概率（通常假设基本事件是统计独立的）决定顶事件发生的概率。因为（相对于表示产品成功运行准则的可靠性框图而言）故障树表示失效准则，所以故障树的分析是通过生成产品的割集完成的。割集是产品的一个二元路径集（a Dual of a Path Set），因为如果所有割集内的部件都失效了，那么产品也就失效了。

因此，路径集建立了产品成功运行的准则，而割集则建立了失效准则。最小割集包含能够使产品失效的最小数量的元素。

1. 割集的生成

决定故障树割集的自顶而下算法从故障树的顶事件开始，在每个低级别门处构建一系列割集，并把这些割集扩展到故障树的每个级别，直到扩展到基本事件级别。假如逻辑门是一个AND（与）门，那么所有的输入事件必须发生，以启动此门。因此在下一级别上，逻辑门可替换为一个输入事件列表。

假如逻辑门是OR（或）门，那么所有建立起来的割集可以分割成若干割集——每个割集包含输入到OR门的事件。例如图11.14所示的故障树，它包含5个逻辑门（G_1 到 G_5）和5个基本事件（A_1 到 A_5）。

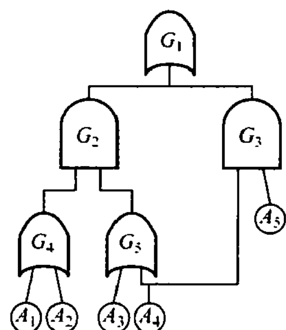


图11.14　用于生成割集的故障树

由此故障树推导出来的割集系列如图11.15所示。自顶向下算法开始于顶门G1，因为 G_1 是一个OR门，所以在扩展中，它由输入 G_2 和 G_3 代替。G_3 是一个AND门，在扩展中被基本事件 $\{A_4, A_5\}$ 代替，$\{A_4, A_5\}$ 是故障树的一个割集。G_2 被扩展成 $\{G_4, G_5\}$，因为它们必须发生以激活 G_2。G_4 的扩展把割集划分成两个集合，因为它是一个具有两个输入事件的OR门：$\{A_1, G_5\}$ 和 $\{A_2, G_5\}$。最后，G_5 的扩展又将以上两个集合分别划分成四个集合，即 $\{A_1, A_3\}$、$\{A_1, A_4\}$、$\{A_2, A_3\}$ 和 $\{A_2, A_4\}$，这些就是故障树的最小割集。

假如逻辑门的扩展是一个 n 中取 k 门，那么它的扩展将是OR门和AND门的一个组

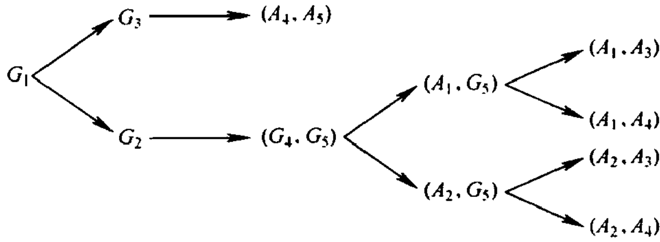


图 11.15　用于确定故障树割集的自顶向下算法

合形式。n 中取 k 门被扩展成输入事件的 c（n，k）组合，此组合可激活门事件。例如某割集的逻辑门 G_x 是一个可以把割集分解成 4 组割集的 4 门中取 3 门，其中的 G_x 就可以替换为输入所选择的 4 种事件组合。

当用这种算法生成割集时，需要对其进行一些简化。假如割集包含了多个相同基本事件，那么就可以排除冗余条目。

假如某割集是另一个割集的子集，那么从长远角度考虑，就可以删除后者。例如割集集合 $\{\{A_1, A_2, A_1, A_3\}, \{A_3, A_4\}, \{A_2, A_3, A_4\}\}$ 就可以简化为 $\{\{A_1, A_2, A_3\}, \{A_3, A_4\}\}$。

2. 包含/排除法

一旦确定了故障树的最小割集，就可以计算出产品失效的概率。割集代表了产品所有可能失效的方式，简单地说，产品的失效概率就是一个或多个割集中的事件将要发生的概率。

$$Pr\{product\ failure\} = Pr\left\{\bigcup_{i=1}^{n} C_i\right\} \tag{11.10}$$

其中，C_i 表示产品的最小割集。因为割集通常是不相交的，所以割集的合并不等于所有单个割集的概率之和。

包含/排除法是计算两个事件相并之后概率的一般性准则。

$$Pr\left\{A \bigcup B\right\} = Pr\{A\} + Pr\{B\} - Pr\{A \Lambda B\} \tag{11.11}$$

其计算方法为

$$Pr\left\{\bigcup_{i=1}^{n} C_i\right\} = \sum_{i=1}^{n} Pr\{C_i\} - \sum_{i<j} Pr\{C_i \Lambda C_j\} + \sum_{i<j<k} Pr\{C_i \Lambda C_j \Lambda C_k\} \mp \cdots \mp Pr\{\mathop{\Lambda}_{i=1}^{n} C_i\} \tag{11.12}$$

公式（11.12）可以精确计算出产品的失效概率。当每个连续割集的总和计算出来后并进行累加，那么所得结果将高估（假如各项是相加的）或者低估（假如各项是相减的）真实的失效概率。因此，只使用公式（11.12）中的一部分就可以计算产品失效概率的边界。

假设有如图 11.14 所示的故障树，它的割集是：$C_1 = \{A_4, A_5\}$，$C_2 = \{A_1, A_3\}$，$C_3 = \{A_1, A_4\}$，$C_4 = \{A_2, A_3\}$ 和 $C_5 = \{A_2, A_4\}$。假设每个基本事件发生的概率是：

$Pr\{A_1\} = q_1 = 0.05$，$q_2 = 0.10$，$q_3 = 0.15$，$q_4 = 0.20$ 和 $q_5 = 0.25$，那么每个割集发

生的概率是：$Pr\{C_1\} = q_4 \times q_5 = 0.05$，$Pr\{C_2\} = 0.0075$，$Pr\{C_3\} = 0.01$，$Pr\{C_4\} = 0.015$，$Pr\{C_5\} = 0.02$。所有割集的概率之和是0.1025，它是产品不可靠度的上边界：

$$0 \leqslant unreliability \leqslant 0.1025 \tag{11.13}$$

公式（11.12）的第二项是所有两个割集的可能组合发生的概率之和，用以上例子来说就是0.015175。用第一项的值中减去这个数值将得到此产品不可靠度的下边界，即0.087325：

$$0.087325 \leqslant unreliability \leqslant 0.1025 \tag{11.14}$$

第三项是三种割集的所有可能组合发生的概率之和，其值为0.0020875，加上它以后将产生一个优化后的上边界：

$$0.087325 \leqslant unreliability \leqslant 0.0894125 \tag{11.15}$$

第四项是四种割集的所有可能组合发生的概率之和，其值为0.0003，减去它以后将产生一个更为严格的下边界：

$$0.0891125 \leqslant unreliability \leqslant 0.0894125 \tag{11.16}$$

假如我们仅需要精确到小数点后三位，那么扩展就可以到此为止，即得到的不可靠度为0.089。如果加上最后一项——五个割集都发生的概率（0.0000375），所得结果就是精确的不可靠度：

$$unreliability = 0.08915 \tag{11.17}$$

11.2 时间相关性

在前一部分中，我们假设部件的失效概率是一个固定值，但在大多数情况下，失效的概率依赖于部件进入运行状态之后时间消逝的总量。在下文的讨论中，将用到以下符号：

$f_i(t)$——部件 i 的失效时间的密度函数；

$F_i(t)$——部件 i 的失效时间的累计分布函数；

$p_i(t)$——部件 i 在 t 时刻的可靠度函数；

$h(t)$——t 时刻的瞬时失效率（故障率）；

MTTF——部件或产品的平均失效时间或平均寿命。

假如部件可靠度和时间的关系已知，则时间因素不会改变用于估计产品可靠度的方法。在任何公式中，$p_i(t)$都可以由 p_i 代替；同样，$q_i(t)$也可以由 q_i 代替。例如用于计算TMR结构部件的可靠度的公式（11.8）就可表示为

$$R_{TMR}(t) = p_V(t) \times [3p_A(t)^2 - 2p_A(t)^3] \tag{11.18}$$

假设在TMR结构中，部件的失效时间服从指数分布，λ_A 是基本部件的失效率参数，λ_V 是表决器的失效率参数。则

$$p_A(t) = e^{-\lambda_A t} \quad p_V(t) = e^{-\lambda_V t} \tag{11.19}$$

将其带入公式（11.18），得：

$$R_{TMR}(t) = e^{-\lambda_V t}(3e^{2\lambda_A t} - 2e^{-3\lambda_A t}) \tag{11.20}$$

进一步假设基本部件以每小时 $\lambda_A = 10^4$ 的速率失效，同时表决器以每小时 $\lambda_A = 10^6$ 的速率失效。作为时间函数的 TMR 结构的可靠度与单个部件可靠度的比较，结果如图 11.16所示。注意：TMR 产品并不总比单个好，它取决于产品使用时间的长度。对于短期运行，TMR 产品比单个部件产品好；但对于长期运行，单个部件更可靠。这种行为的发生与单个部件的可靠度和表决器的可靠度无关。

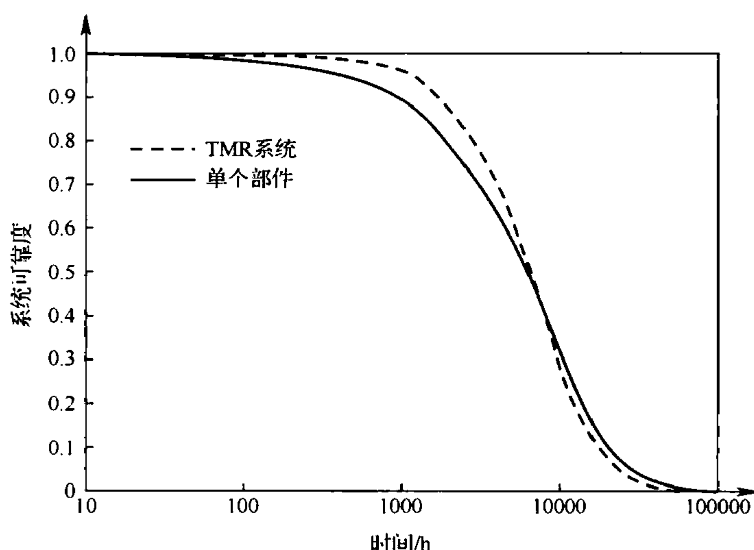


图 11.16　TMR 可靠性与单个部件可靠性的比较

11.2.1　平均失效时间

对于那些失效概率与时间相关的产品，一个重要的参数是平均失效时间（Mean Time To Failure，MTTF）或者平均寿命（Average Lifetime）。假如 $f(t)$ 代表（有冗余部件系统或无冗余部件系统）失效时间密度，那么平均寿命可由式（11.21）计算：

$$\text{MTTF} = \int_0^\infty x f(x)\,\mathrm{d}x \tag{11.21}$$

例如单个部件的失效时间服从参数为 λ_A 的指数分布，那么此部件的平均寿命是

$$\text{MTTF}_A = \int_0^\infty x\lambda_A \mathrm{e}^{-\lambda_A x}\,\mathrm{d}x = \frac{1}{\lambda_A} \tag{11.22}$$

产品的平均失效时间可由可靠度公式计算：

$$\text{MTTF}_S = \int_0^\infty R_S(t)\,\mathrm{d}t \tag{11.23}$$

对于先前介绍的 TMR 产品，单个部件的失效率为 λ_A，表决器的失效率为 λ_V，则其 MTTF 为

$$\text{MTTF}_{\text{TMP}} = \int_0^\infty \mathrm{e}^{-\lambda_V t}(3\mathrm{e}^{-2\lambda_A t} - 2\mathrm{e}^{-3\lambda_A t}) = \frac{3}{\lambda_V + 2\lambda_A} - \frac{2}{\lambda_V + 3\lambda_A} \tag{11.24}$$

比较公式（11.22）和公式（11.24），即使乐观地假设表决器绝对不会失效（即 $\lambda_V=0$），单个部件的MTTF也要比TMR产品的MTTF好。这表明MTTF并不是用于比较产品可靠度最好的尺度。回忆一下TMR产品，当 t 很小时，其可靠性最高；但 t 很大时，其可靠性却较低。

11.2.2　故障率

另一个测量时间关联的失效概率的量是故障率 $h(t)$。故障率或瞬时失效率指产品或部件至少存活到 t 时刻的失效率。假定产品可以存活到 t 时刻，$h(t)\Delta t$ 表示部件在时间间隔 $(t, t+\Delta t)$ 内失效的条件概率：

$$h(t)=\lim_{\tau\to 0}\frac{1}{\tau}\frac{F(t+\tau)-F(t)}{R(t)}=\frac{f(t)}{R(t)} \tag{11.25}$$

对于失效时间为指数分布的部件，其失效率是常数：

$$h_i(t)=\frac{f(t)}{R(t)}=\frac{\lambda_i e^{-\lambda_i t}}{e^{-\lambda_i t}}=\lambda_i \tag{11.26}$$

只有失效时间服从指数分布，部件才拥有恒定失效率。假如部件的失效时间均匀分布在 a 和 b 之间，那么失效率将随着 t 的增长而接近于 b：

$$h_i(t)=\frac{f(t)}{R(t)}=\frac{\frac{1}{b-a}}{\frac{b-t}{b-a}}\quad t\in[a,b] \tag{11.27}$$

失效函数的确定导出了另定义可靠度的另一种方式。假设 $R(0)=1$，对公式（11.25）两边求积分，可得：

$$\int_0^t h(x)\mathrm{d}x=\int_0^t\frac{f(x)}{R(x)}\mathrm{d}x=\int_0^t\frac{-R'(x)}{R(x)}\mathrm{d}x=-\ln(R(t)) \tag{11.28}$$

则

$$R(t)=e^{-\int_0^t h(x)\mathrm{d}x} \tag{11.29}$$

11.3　动态冗余——Markov模型

使用动态冗余的产品可以在部件失效之后自动重组。例如冗余单元可能在需要其运行之前都保持无动力状态，当前主要单元失效后，它才转入到运行状态。图11.7是一个具有备用冗余的产品。部件 A_P 是初始运行的主要部件。当 A_P 失效时，备用单元 A_S 转入运行状态。动态冗余的另一个例子是TMR/单一产品，在此类产品中，当第一次失效出现时，TMR产品将重组。在标准的TMR产品中，只有三分之一的部件失效，剩余两个部件的一个也失效时，产品才会失效。与TMR/单一产品一样，假如动态冗余产品中剩

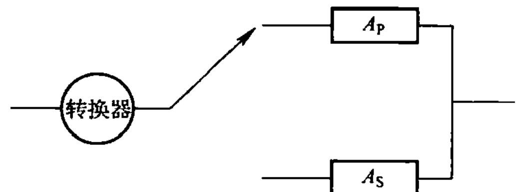


图11.17　备用冗余

余两个部件中的一个被弃用，产品仍将在下一个部件失效的时候发生失效，但其发生的概率要低于前一种。也就是说，一个部件失效的概率高于两个有效部件中只有一个可用时的失效概率。在每次失效之后，TMR 产品能够通过移除一个好的部件实现动态重组（这样就可以保持奇数个有效部件）。

分析这种能够重组的产品比分析静态冗余产品要复杂得多，因为失效准则取决于失效发生的顺序，而不仅仅取决于部件的组合形式。例如在图 11.17 所示的备用冗余产品中，只要主要单元可运行，那么产品就可运行。

假如主要单元失效，只要转换器还没有失效，那么产品就可以运行，因为备用单元可以转入到工作状态。假如转换器在起动备用单元后失效，那么产品仍旧可以运行（假设转换器失效意味着转换器的状态不能再改变）。主要单元失效和转换器失效的组合有时也会引起产品失效，但并不总是这样，产品是否失效取决于失效发生的顺序。

假设所有部件的失效时间服从指数分布，那么就使用 Markov 链来评估动态冗余产品的可靠度。形象地说，一个 Markov 链由表示产品状态的环组成，这些环由代表改变产品状态的事件（通常是失效）的与门符号连接起来。而与门符号的标签是事件的发生率。

Markov 链生成了一组线性的常微分方程。设 $p_i(t)$ 是 Markov 链在 t 时刻处于状态 i 的概率。产品的可靠度计算为 $p_i(t)$ 处于可运行状态概率的总和：

$$R(t) = \sum_{i \in \text{可运行状态}} p_i(t) \tag{11.30}$$

在以下内容中，我们将用 Markov 链分析几个典型的产品。

11.3.1 备用冗余

我们讨论的第一个备用冗余产品如图 11.17 所示，假设主要单元和备用单元（曾经是工作单元）的失效率是 λ_P，开关的失效率为 λ_S。此产品的 Markov 链表示如图 11.18 所示。

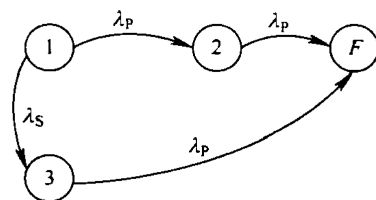


图 11.18 冗余系统的 Markov 链

状态 1 表示产品的初始状态，在此状态下，主要部件处于运行状态。初始状态产生两个事件，每个事件都导致不同的后续状态。假如主要单元首先失效，那么产品就会转入状态 2。在状态 2 中，备用单元已经转换到运行状态。假如在状态 1 中，转换器首先失效，那么主要单元仍旧是可运行的，产品转向状态 3。

在状态 2 中，当工作单元（以前是备用单元）失效时，产品就会转入失效状态。假如转换器从状态 2 开始失效，那么产品的可靠度不会受到影响，因此，此事件不包括在 Markov 链之内。在状态 3 中，当主要单元失效时，产品也失效，因为备用单元无法转换到工作状态。产品的可靠度由产品处于状态 1、2、3 的概率决定，产品的不可靠度则取决于产品处于状态 F 的概率。

与图 11.18 所示的 Markov 链相关的公式如下：

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}p_1(t) &= -(\lambda_P+\lambda_S)p_1(t)\\
\frac{\mathrm{d}}{\mathrm{d}t}p_2(t) &= \lambda_P p_1(t)-\lambda_P p_2(t)\\
\frac{\mathrm{d}}{\mathrm{d}t}p_3(t) &= \lambda_S p_1(t)-\lambda_P p_3(t)\\
\frac{\mathrm{d}}{\mathrm{d}t}p_F(t) &= \lambda_P p_2(t)+\lambda_P p_3(t)
\end{aligned} \tag{11.31}
$$

这组方程很容易由 Laplace 变换求解。第一个状态是初始状态——即 $p_1(0)=1$，$p_i(t)=0(i\neq 1)$。对方程两边进行 Laplace 变换，将会得到以下方程组，其中，$L_i(s)$ 表示 $p_i(t)$ 的 Laplace 变换：

$$
\begin{aligned}
sL_1(s)-1 &= -(\lambda_P+\lambda_S)L_1(s)\\
sL_2(s) &= \lambda_P L_1(s)-\lambda_P L_2(s)\\
sL_3(s) &= \lambda_S L_1(s)-\lambda_P L_3(s)\\
sL_F(s) &= \lambda_P L_2(s)-\lambda_P L_3(s)
\end{aligned} \tag{11.32}
$$

用此方程组求解 L_i，可得：

$$
\begin{aligned}
L_1(s) &= \frac{1}{s+\lambda_P+\lambda_S}\\
L_2(s) &= \frac{\lambda_P}{\lambda_S}\left(\frac{1}{s+\lambda_P}-\frac{1}{s+\lambda_P+\lambda_S}\right)\\
L_3(s) &= \frac{1}{s+\lambda_P}-\frac{1}{s+\lambda_P+\lambda_S}\\
L_F(s) &= \frac{1}{s}-\frac{\lambda_P+\lambda_S}{\lambda_S}\frac{1}{s+\lambda_P}+\frac{\lambda_P}{\lambda_S}\frac{1}{s+\lambda_P+\lambda_S}
\end{aligned} \tag{11.33}
$$

应用 Laplace 反变换，可得出产品的状态概率（表 11.1 列出一些有用的 Laplace 变换）：

表 11.1　一些有用的 Laplace 变换

$L_F(s)$	$f(t)$，$t>0$
$\frac{c}{s}$	常数 C
$\frac{1}{s+a}$	e^{-at}
$\frac{1}{(s+a)(s+b)}$	$\frac{1}{b-a}(e^{-at}-e^{-bt})$
$\frac{1}{(s+a)(s+b)(s+c)}$	$\frac{e^{-at}}{(b-a)(c-a)}+\frac{e^{-bt}}{(a-b)(c-b)}+\frac{e^{-ct}}{(a-c)(b-c)}$
$\frac{1}{(s+a)(s+b)(s+c)(s+d)}$	$\frac{e^{-at}}{(b-a)(c-a)(d-a)}+\frac{e^{-bt}}{(a-b)(c-b)(d-b)}+\frac{e^{-ct}}{(a-c)(b-c)(d-b)}+\frac{e^{-dt}}{(a-d)(b-d)(c-d)}$

$$
\begin{aligned}
P_1(t) &= e^{-(\lambda_P+\lambda_S)t} \\
P_2(t) &= \frac{\lambda_P}{\lambda_S}[e^{-\lambda_P t} - e^{-(\lambda_P+\lambda_S)t}] \\
P_3(t) &= e^{-\lambda_P t} - e^{-(\lambda_P+\lambda_S)t} \\
P_F(t) &= 1 - \frac{\lambda_P+\lambda_S}{\lambda_S}e^{-\lambda_P t} + \frac{\lambda_P}{\lambda_S}e^{-(\lambda_P+\lambda_S)t}
\end{aligned} \tag{11.34}
$$

因此，图 11.17 所示的备用冗余产品的可靠度将可由式（11.35）计算：

$$
R(t) = \frac{\lambda_P+\lambda_S}{\lambda_S}e^{-\lambda_P t} - \frac{\lambda_P}{\lambda_S}e^{-(\lambda_P+\lambda_S)t} \tag{11.35}
$$

对于上一个备用冗余产品，我们假定备用单元是部分通电的，那么备用单元就可能在转入到运行状态之前失效。相对于冷备用（Cold Spare）来说，这种备用称为温备用（Warm Spare）。温备用在工作之前不能失效，而热备用（Hot Spare）则完全工作，并和主要单元拥有相同的失效率。具有温备用的备用冗余产品的 Markov 模型如图 11.19 所示。此 Markov 链与图 11.18所示的链相似。

但是从状态 1 到状态 3 的转换速率从 λ_S（表示转换器失效）增长到了 $\lambda_S+\lambda_W$（表示转换器或备用失效）。在具有温备用的产品中，如果在备用单元在主要单元失效前失效（失效率为 λ_W），那么这种现象与转换器失效相同，因为主要单元失效时，产品也会失效。具有温备用的产品的可靠度由式（11.36）计算：

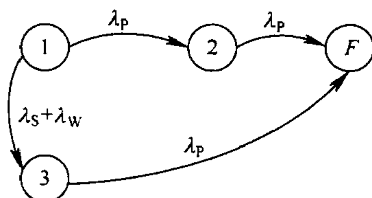


图 11.19　具有温备用系统的 Markov 模型

$$
R(t) = \frac{\lambda_P+\lambda_S+\lambda_W}{\lambda_S+\lambda_W}e^{-\lambda_P t} - \frac{\lambda_P}{\lambda_S+\lambda_W}e^{-(\lambda_P+\lambda_S+\lambda_W)t} \tag{11.36}
$$

其中，$(\lambda_S+\lambda_W)$ 代替了方程式（11.35）中的 λ_S。

11.3.2　TMR/单一系统

TMR/单一系统是我们举的第二个例子，TMR 系统可以在第一次失效出现时重组成单个部件。此类系统的 Markov 链表示如图 11.20 所示，其 Laplace 变换方程如下：

$$
\begin{aligned}
sL_1(s) - 1 &= -(3\lambda_A+\lambda_V)L_1(s) \\
sL_2(s) &= 3\lambda_A L_1(s) - (\lambda_A+\lambda_V)L_2(s) \\
sL_F(s) &= (\lambda_A+\lambda_V)L_2(s) + \lambda_V L_1(s)
\end{aligned} \tag{11.37}
$$

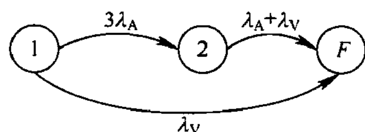


图 11.20　TMR/单一系统的 Markov 模型

系统的状态概率为

$$
\begin{aligned}
p_1(t) &= e^{-(3\lambda_A+\lambda_V)t} \\
p_2(t) &= \frac{3}{2}(e^{-(\lambda_A+\lambda_V)t} - e^{-(3\lambda_A+\lambda_V)t})
\end{aligned} \tag{11.38}
$$

系统的可靠度计算为

$$
R_{\text{TMR/单一系统}}(t) = \frac{3}{2}e^{-(\lambda_A+\lambda_V)t} - \frac{1}{2}e^{-(3\lambda_A+\lambda_V)t} \tag{11.39}
$$

图11.21比较了TMR/单一系统、标准TMR系统和单独的非冗余部件之间时间相关可靠度。与前面描述的一样，我们假设部件以每小时$\lambda_A=10^{-4}$的速率失效，表决器以每小时$\lambda_V=10^{-6}$的速率失效。重组系统始终比标准TMR系统和单个部件系统的可靠度都高，但如果表决器的可靠度不高，那么TMR/单一系统的可靠度就没有单个部件系统的可靠度高。表决器本身的可靠度低于单个部件的可靠度（$\lambda_V>\lambda_A$）清晰地印证了这个观点。因为从可靠性的观点来看，表决器通常与三个部件串联相连，而三个部件的可靠度绝对不会比单个部件的可靠度高。在这种情况下，表决产品的可靠度决不会比单个部件的可靠度高。

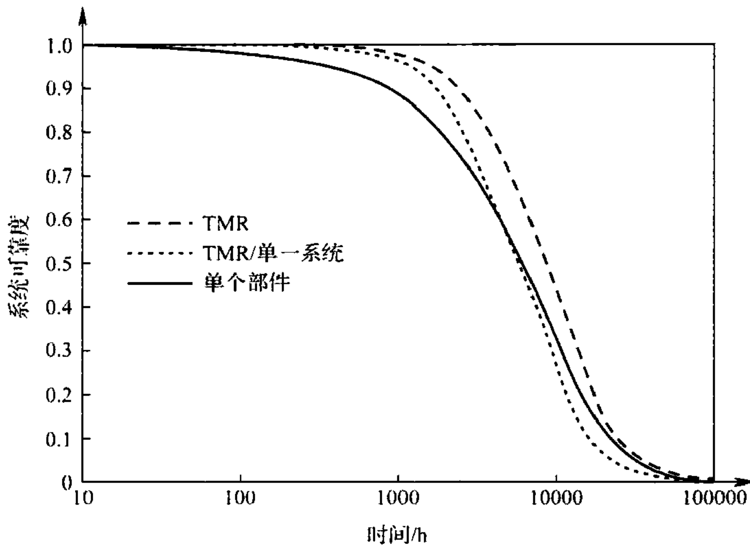


图11.21 TMR、TMR/单一系统和单个部件系统的可靠度比较

TMR/单一系统的MTTF可由式（11.40）计算：

$$MTTF_{TMR/单一系统}=\int_0^\infty\left(\frac{3}{2}e^{-(\lambda_A+\lambda_V)t}-\frac{1}{2}e^{-(3\lambda_A+\lambda_V)t}\right)dt=\frac{\frac{3}{2}}{\lambda_A+\lambda_V}-\frac{\frac{1}{2}}{3\lambda_A+\lambda_V}\quad(11.40)$$

为了比较TMR/单一系统、标准TMR系统以及单个部件系统的MTTF，我们假定表决器不会失效（即$\lambda_V=0$），则平均失效时间为

$$\mathrm{MTTF}_{TMR/单一系统}=\frac{3}{(2\lambda_A)}-\frac{1}{6\lambda_A}=\frac{4}{3\lambda_A}$$

$$\mathrm{MTTF}_{TMR}=\frac{3}{2\lambda_A}-\frac{2}{3\lambda_A}=\frac{5}{6\lambda_A}\quad(11.41)$$

$$\mathrm{MTTF}_A=\frac{1}{\lambda_A}$$

假设表决器的表现非常良好，则TMR/单一系统的MTTF不仅好于标准TMR系统的MTTF，也好于单个部件系统的MTTF。

11.3.3　可修复产品

假如一个产品在失效之后可以修复，那么我们将无法对其可靠度进行适当的度量，因为此类产品的可靠度的定义为时间间隔 $[0, t]$ 内产品不失效的概率。实际上，可靠度并不能反映产品是否是可修复的。一种更有效的用于测量可修复产品效能的量度是可用性 $[A(t)]$，它的定义为产品在 t 时刻运行的概率。稳态可用性（Steady-State Availability）（A）是产品可运行的时间相关概率。$A(t)$ 和 A 都认为产品可处于 UP 状态或 DOWN 状态，因此它反映了产品瞬时或长期可运行的概率。

可修复产品可以建模为组合模型或 Markov 模型，这取决于对修复设备的假设。假如我们认为产品的部件的修复是相互独立的，且组合模型适用于评估产品的可靠性（在没有修复的情况下），那么同样的模型就可用于评估产品的可用性。但假如可用的修复人员少于部件，那么就要使用 Markov 模型对产品进行建模。下一部分中，我们将分别讨论这两种情况。

1. 独立修复

假如失效部件的修复过程全部独立——也就是说，存在充足的修复设备，没有等待和优先权之类的问题——那么组合模型既可以用来评估可用性，也可以用来评估可靠性。例如在一个简单的 TMR 产品中，其部件和表决器都是可以修复的。此时，TMR 产品的可靠度由式（11.42）计算：

$$R_{\mathrm{TMR}} = p_{\mathrm{V}} \times [p_{\mathrm{A}}^3 + 3p_{\mathrm{A}}^2(1 - p_{\mathrm{A}})] = p_{\mathrm{V}} \times (3p_{\mathrm{A}}^2 - 2p_{\mathrm{A}}^3) \tag{11.42}$$

假如 av_{A} 表示部件 A 的运行时间中的长期运行部分，av_{V} 表示表决器的运行时间中的长期运行部分，那么将上式中的 p 替换为 av，就可以计算处 TMR 产品的可用性：

$$A_{\mathrm{TMR}} = av_{\mathrm{V}} \times (3av_{\mathrm{A}}^2 - 2av_{\mathrm{A}}^3) \tag{11.43}$$

TMR 产品的时间相关可用性可以通过 $av_{\mathrm{A}}(t)$ 和 $av_{\mathrm{V}}(t)$ 的时间相关可用性来计算：

$$A_{\mathrm{TMR}}(t) = av_{\mathrm{V}}(t) \times [3av_{\mathrm{A}}^3(t)] \tag{11.44}$$

单个部件的时间相关可用性可以用平均失效时间（MTTF）和平均修复时间（MTTR）确定：

$$av_{\mathrm{A}} = \frac{\mathrm{MTTF}}{\mathrm{MTTF} + \mathrm{MTTR}} \tag{11.45}$$

在失效时间 λ_{A} 和修复时间 μ_{A} 都服从指数分布的情况下，部件的可用性由式（11.46）计算：

$$av_{\mathrm{A}} = \frac{\mathrm{MTTF}}{\mathrm{MTTF} + \mathrm{MTTR}} = \frac{\frac{1}{\lambda_{\mathrm{A}}}}{\frac{1}{\lambda_{\mathrm{A}}} + \frac{1}{\mu_{\mathrm{A}}}} = \frac{\mu_{\mathrm{A}}}{\mu_{\mathrm{A}} + \lambda_{\mathrm{A}}} \tag{11.46}$$

在同样的情况下（失效时间和修复时间都服从指数分布），部件的时间相关可用性可以用 Markov 模型分析来确定。用简单的 Markov 模型表示的单个部件的运行和失效状态之间的转变如图 11.22 所示。如前所述，产品的状态概率可以通过使用 Laplace 的变

换确定。部件在 t 时刻处于状态 UP 的概率就是其可用性：

$$sL_{\mathrm{UP}}(s)-1=-\lambda_{\mathrm{A}}L_{\mathrm{UP}}+\mu_{\mathrm{A}}L_{\mathrm{DOWN}}=-\mu_{\mathrm{A}}L_{\mathrm{DOWN}}+\lambda_{\mathrm{A}}L_{\mathrm{UP}} \tag{11.47}$$

$$L_{\mathrm{UP}}=\frac{1+\mu_{\mathrm{A}}L_{\mathrm{DOWN}}}{s+\lambda_{\mathrm{A}}}$$

$$L_{\mathrm{DOWN}}=\frac{\lambda_{\mathrm{A}}L_{\mathrm{UP}}}{s+\mu_{\mathrm{A}}} \tag{11.48}$$

图 11.22　单个可修复部件的 Markov 模型

将 L_{DOWN} 的表达式带入 L_{UP} 的表达式就可以求出 L_{UP}：

$$L_{\mathrm{UP}}=\frac{s+\mu_{\mathrm{A}}}{s^2+s\mu_{\mathrm{A}}+s\lambda_{\mathrm{A}}}=\frac{1}{s+\mu_{\mathrm{A}}+\lambda_{\mathrm{A}}}+\frac{\mu_{\mathrm{A}}}{s(s+\lambda_{\mathrm{A}}+\mu_{\mathrm{A}})} \tag{11.49}$$

通过 Laplace 反变换，可以求出部件的状态概率和可用性：

$$A(t)=P_{\mathrm{UP}}(t)=\frac{\mu_{\mathrm{A}}}{\mu_{\mathrm{A}}+\lambda_{\mathrm{A}}}+\frac{\lambda_{\mathrm{A}}}{\mu_{\mathrm{A}}+\lambda_{\mathrm{A}}}\mathrm{e}^{-(\lambda_{\mathrm{A}}+\mu_{\mathrm{A}})t} \tag{11.50}$$

对参数 t 接近于无穷大时求极限，就可得到时间相关（稳态）可用性。

2. 非独立修复

假如不是每个部件都配有独立的修复人员，那么修复过程将不独立，因为部件可能需要等待才能得到修复。

假如某 TMR/单一产品有两个可用的修复技师：一个是修复表决器的，另一个是修复冗余部件的。此产品的 Markov 模型如图 11.23 所示。图中，μ_{A} 是单个失效部件的修复率，μ_{V} 是表决器的修复率（假如这两个技师都可以修复单个部件，那么部件就可以同时修复，产品从状态 FA 转换到状态 2 的速率将是 $2\mu_{\mathrm{A}}$）。产品的稳态可用性是产品处于状态 3 或者 2 的稳态概率。

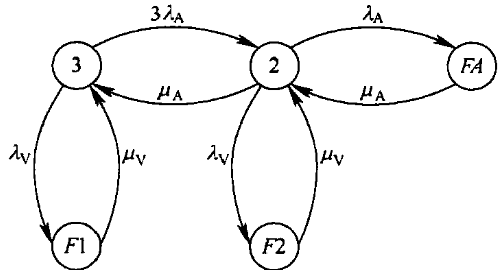


图 11.23　可独立修复的可修复 TMR/单一系统

Markov 链的稳态概率很容易通过平衡方程计算得到。在稳定状态下，产品的状态概率是不变的（$\mathrm{d}P/\mathrm{d}t=0$），且转入状态的总速率等于转出的总速率。

相等的转入和转出速率将产生一组平衡代数式，这些代数式可用于计算产品的状态概率。图 11.23 所示的 Markov 链的平衡方程为

$$\begin{aligned}(3\lambda_{\mathrm{A}}+\lambda_{\mathrm{V}})p_3&=\mu_{\mathrm{A}}p_2+\mu_{\mathrm{V}}p_{\mathrm{F_1}}\\(\lambda_{\mathrm{A}}+\mu_{\mathrm{A}}+\lambda_{\mathrm{V}})p_2&=3\lambda_{\mathrm{A}}p_3+\mu_{\mathrm{A}}F_{\mathrm{A}}+\mu_{\mathrm{V}}F_2\\\lambda_{\mathrm{V}}p_3&=\mu_{\mathrm{V}}p_{F1}\\\lambda_{\mathrm{V}}p_2&=\mu_{\mathrm{V}}p_{F2}\\\lambda_{\mathrm{A}}p_2&=\mu_{\mathrm{A}}p_{\mathrm{FA}}\end{aligned} \tag{11.51}$$

求解以上方程组的 p_3：

$$
\begin{aligned}
p_3 &= p_3 \\
p_2 &= 3\frac{\lambda_A}{\mu_A}p_3 \\
p_{F1} &= \frac{\lambda_V}{\mu_V}p_3 \\
p_{F2} &= \frac{3\lambda_A\lambda_V}{\mu_A\mu_V}p_3 \\
p_{FA} &= \frac{3\lambda_A^2}{\mu_A^2}p_3
\end{aligned}
\tag{11.52}
$$

将已知条件 $p_3+p_2+p_{F1}+p_{F2}+p_{FA}=1$ 代入方程，解出 p_3：

$$
p_3 = \frac{\mu_A^2\mu_V}{\mu_A^2\mu_V+3\lambda_A\mu_A\mu_V+\lambda_V\mu_A^2+3\lambda_A\lambda_V\mu_A+3\lambda_A^2\mu_V}
\tag{11.53}
$$

从式（11.53）中也能计算出产品的另一个状态概率。产品的可用性可以用 $A=p_2+p_3$ 计算，因为这两种状态表示了产品的运行结构：

$$
A = \frac{\mu_A^2\mu_V+3\lambda_A\mu_A\mu_V}{\mu_A^2\mu_V+3\lambda_A\mu_A\mu_V+\lambda_V\mu_A^2+3\lambda_A\lambda_V\mu_A+3\lambda_A^2\mu_V}
\tag{11.54}
$$

11.4　关联失效

在前面的内容中，我们假定产品的部件失效是相互独立的，但在某些情况下，这种假设是不成立的。当一个部件的失效引起其他部件失效时，就会发生共模失效（Common-Mode Failure）。例如与多个部件相连的动力供给失效，那么与之相连的部件也就会失效。当一个部件的失效能增加其他部件失效的概率时，就会出现另一种类型的关联失效。这种现象存在于某些部件共同承担载荷的时候，且某个部件的失效将会给剩余部件带来额外载荷。关联失效的第三种类型与那些具有多种失效方式的部件有关，例如二极管可能由于短路或者断路而失效。这种部件的失效称作多模失效模式（Multiple Failure Modes）或者多模失效（Multimode Failures）。

11.4.1　共模失效

分析受共模失效影响的产品要相对简单一些，组合模型或者 Markov 模型都可以用来对此类产品进行分析。如图 11.2 所示的简单冗余产品，假设部件 A_1 和 B_2 都连接到动力供给 C_1，而部件 A_2 和 B_1 都连接到另一个动力供给 C_2。假如其中一个动力供给失效，那么与之相连接的部件将不能运行，也就是说部件失效了。如图 11.24 所示一样，用串联中某部件及其动力供给代替可靠性框图中的这个部件时，将会影响框图中动力供给的表示。因为部件不仅出现于一条路径上，所以用上述方法绘制的框图表示的不再是简单的串并联产品。在产品的故障树中，每个表示部件的基本事件都可以由 *OR* 门、部件和其动力供给替代。例如图 11.13 所示系统的故障树就可变成图 11.25 所示的故障树。

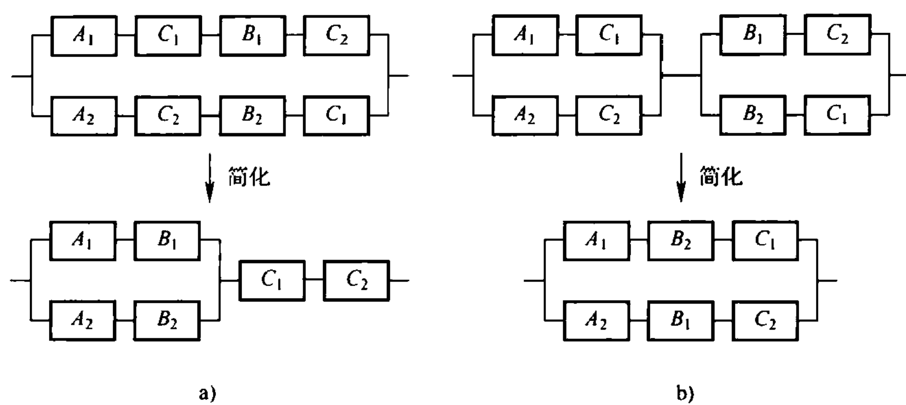


图 11.24　包括动力供给的简单冗余模型
a）系统级冗余　b）部件级冗余

当使用动态冗余或者产品的修复具有非独立性时，用Markov模型为共模失效建模就要相对简单一些。从已知初始状态开始，当共模部件发生失效，就为结果状态加上一个与门符号，其标签是共模部件的失效率。例如在TMR/单一产品中，如果三个冗余部件连接到同一动力供给，那么就为初始状态（状态1）和失效状态（状态F）之间加上一个与门符号。在这种情况下，动力供给失效将致使冗余变得毫无用处。实际上，从可靠性观点来说，动力供给失效与表决器失效具有相同效果。

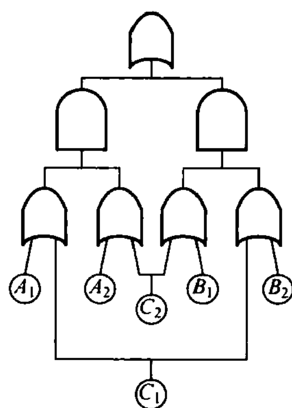


图 11.25　包括动力供给的故障树模型

11.4.2　关联失效率

在一些产品中，部件的失效概率会改变其他部件失效状况。例如有些产品中可能会有两个冗余部件共享同一载荷的情况。当其中一个部件失效时，另一个部件将工作得相当吃力或者其失效概率会增加。Markov模型可以轻松应对这种情况，因为与门符号处的状态转换率会因初始状态的不同而有所差异。以图11.26所示产品的可靠性框图为例。假如冗余部件A_1和A_2共享同一载荷，部件B_1和B_2也共享另一个载荷；当冗余部件组都运行时，它们分别以λ_{A1}和λ_{B1}的速率失效。在A组部件中的一个发生失效之后，系统仍会以λ_{A2}的速率失效。同样，当只剩下B组部件运行时，系统的失效率仍是λ_{B2}。图11.2所示为此产品的Markov模型，图中的运行状态的标签是一对用于表示A和C组有多少剩余部件的有序整数。假如部件失效率独立于处于运行状态的部件数，那么λ_{A1}等于λ_{A2}，λ_{B1}等于λ_{B2}。在这种情况下，使用故障树或可靠性框图来分析产品就稍微简单一些。

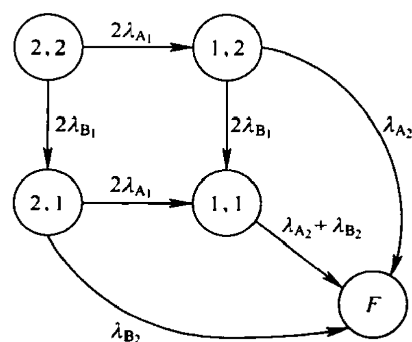


图 11.26　拥有共享载荷部件系统的Markov模型

11.4.3　多模失效

某些部件有多种失效方式。也就是说，与我们目前的假设相比，这些部件至少有两种可能的失效方式。例如双发动机飞行器上的发动机可能丧失一半动力或者完全丧失动力。假设只要有一个发动机具有完全的动力或者两个发动机各有一半动力，飞行器就可以安全着陆，此产品的可靠性框图如图 11.27 所示。图中，Ai_F 表示发动机 i 具有完全动力，Ai_H 表示发动机 i 只具有一半动力。因为框图中的部件是非独立的，所以部件 i 的状态是互斥的。

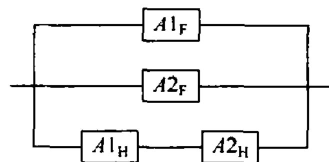


图 11.27　多重失效模式系统的可靠性框图

发动机不可能同时处于全动力或者半动力状态。在扩展路径集时，应该考虑可靠性框图中部件的独立性。此产品的最小路径集是 $\{A1_F\}$、$\{A2_F\}$ 和 $\{A1_H, A2_H\}$。使用包含/排除算法可以求出产品的可靠度：

$$\begin{aligned}R &= Prob(A1_F)+Prob(A2_F)+Prob(A1_H\Lambda A2_H)-Prob(A1_F\Lambda A2_F)-\\&Prob(A1_F\Lambda A1_H\Lambda A2_H)-Prob(A2_F\Lambda A1_H\Lambda A2_H)+Prob(A1_F\Lambda A2_F\Lambda A1_H\Lambda A2_H)\\&=2Prob(A_F)+Prob^2(A_H)-Prob^2(A_F)\end{aligned}\tag{11.55}$$

其中，我们假设统计中的两个发动机是相同的。式（11.55）中最后三项为零，因为部件不可能同时处于两种状态。

假如 A_F、A_H 和 A_0 的概率是已知的，且 $Prob(A_F)+Prob(A_H)+Prob(A_0)=1$，我们可以将其直接代入方程（11.55）中。

假设状态概率不是固定的，全动力发动机失效为半动力时的速率为 λ_{FH}，全动力发动机失效为 0 动力的速率为 λ_{F0}，半动力发动机失效为 0 动力的速率为 λ_{H0}。我们可以在 Markov 模型中（见图 11.28）使用这些速率来求解公式（11.55）中的未知项（此后公式将变得与时间相关）。

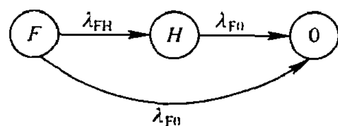


图 11.28　单个发动机多模失效的 Markov 模型

图 11.28 所示 Markov 链的解是部件处于以下三种状态的概率：

$$\begin{aligned}&Porb[A_F(t)]=P_F(t)=e^{-(\lambda_{FH}+\lambda_{F0})t}\\&Porb[A_H(t)]=P_H(t)=\frac{\lambda_{FH}}{\lambda_{H0}-\lambda_{FH}-\lambda_{F0}}[e^{-(\lambda_{FH}+\lambda_{F0})t}-e^{-\lambda_{H0}t}]\\&Porb[A_0(t)]=P_0(t)=1-\frac{\lambda_{H0}-2\lambda_{FH}-\lambda_{F0}}{\lambda_{H0}-\lambda_{FH}-\lambda_{F0}}e^{-(\lambda_{FH}+\lambda_{F0})t}+\frac{\lambda_{FH}}{\lambda_{H0}-\lambda_{FH}-\lambda_{F0}}e^{-\lambda_{H0}t}\end{aligned}\tag{11.56}$$

使用公式（11.55），可求解出飞行器成功飞行的概率：

$$\begin{aligned}R(t)&=2Prob[A_F(t)]+Prob^2[A_H(t)]-Prob^2[A_F(t)]\\&=2e^{-(\lambda_{FH}+\lambda_{F0})t}+\frac{\lambda_{FH}^2}{(\lambda_{H0}-\lambda_{FH}-\lambda_{F0})^2}[e^{-(\lambda_{FH}+\lambda_{F0})t}-e^{-\lambda_{H0}t}]^2-e^{-2(\lambda_{FH}+\lambda_{F0})t}\end{aligned}\tag{11.57}$$

然而，假如组合模型不适合用来为此产品建模，那么我们可以把多重失效模式融入到 Markov 模型中来解决此问题。图 11.29 展示了具有两个发动机产品的 Markov 模型

在状态 $2F$ 中，两个发动机处于全动力状态；在状态 FH 中，一个发动机处于全动力状态，另一个处于是半动力状态；在状态 $2H$ 中，两个发动机都处于半动力状态；在状态 $F0$ 中，一个发动机处于全动力状态另一个处于0动力状态；在状态0中，两个发动机都没有动力，产品失效。对公式展开、求解并进行 Laplace 反变换，就可求得每一个运行状态的概率。

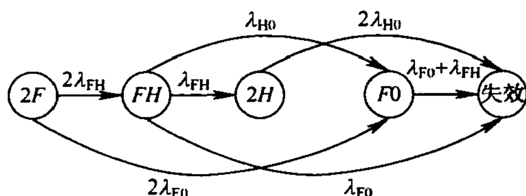


图 11.29 具有多重失效模式的双发动机系统的 Markov 模型

图 11.29 所示的 Markov 链中的运行状态 Laplace 变换为

$$
\begin{aligned}
L_{2F}(s) &= \frac{1}{s+2\lambda_{FH}+2\lambda_{F0}} \\
sL_{FH}(s) &= 2\lambda_{FH}L_{2F}(s)-(\lambda_{FH}+\lambda_{H0}+\lambda_{F0})L_{FH}(s) \\
L_{FH}(s) &= \frac{2\lambda_{FH}}{(s+2\lambda_{FH}+2\lambda_{F0})(s+\lambda_{FH}+\lambda_{H0}+\lambda_{F0})} \\
sL_{2H}(s) &= \lambda_{FH}L_{FH}(s)-2\lambda_{H0}L_{2H}(s) \\
L_{2H}(s) &= \frac{2\lambda_{FH}^2}{(s+2\lambda_{FH}+2\lambda_{F0})(s+\lambda_{FH}+\lambda_{H0}+\lambda_{F0})(s+2\lambda_{H0})} \\
sL_{F0}(s) &= \frac{2\lambda_{F0}}{(s+\lambda_{FH}+\lambda_{F0})(s+2\lambda_{FH}+2\lambda_{F0})}+ \\
&\quad \frac{2\lambda_{FH}\lambda_{H0}}{(s+2\lambda_{FH}+2\lambda_{F0})(s+\lambda_{FH}+\lambda_{H0}+\lambda_{F0})(s+\lambda_{FH}+\lambda_{F0})}
\end{aligned}
\tag{11.58}
$$

通过 Laplace 反变换，可以得出状态概率：

$$
\begin{aligned}
P_{2F}(t) &= e^{-2(\lambda_{FH}+\lambda_{F0})t} \\
P_{FH}(t) &= \frac{2\lambda_{FH}}{\lambda_{H0}-\lambda_{FH}-\lambda_{F0}}\left[e^{-2(\lambda_{FH}+\lambda_{F0})t}-e^{-(\lambda_{FH}+\lambda_{H0}+\lambda_{F0})t}\right] \\
P_{2H}(t) &= \frac{\lambda_{FH}^2}{(\lambda_{H0}-\lambda_{FH}-\lambda_{F0})^2}\left[e^{-2(\lambda_{FH}+\lambda_{F0})t}+2e^{-(\lambda_{FH}+\lambda_{H0}+\lambda_{F0})t}+e^{-2\lambda_{H0}t}\right] \\
P_{F0}(t) &= \frac{2\lambda_{F0}}{\lambda_{FH}+\lambda_{F0}}\left[e^{-(\lambda_{FH}+\lambda_{F0})t}-e^{-2(\lambda_{FH}+\lambda_{F0})t}\right]- \\
&\quad 2\lambda_{H0}\lambda_{FH}\left[\frac{e^{-2(\lambda_{FH}+\lambda_{F0})t}}{(\lambda_{H0}-\lambda_{FH}-\lambda_{F0})(\lambda_{FH}+\lambda_{F0})}+\frac{e^{-(\lambda_{FH}+\lambda_{H0}+\lambda_{F0})t}}{\lambda_{H0}(\lambda_{H0}-\lambda_{FH}-\lambda_{F0})}-\frac{e^{-(\lambda_{FH}+\lambda_{F0})t}}{\lambda_{H0}(\lambda_{FH}+\lambda_{F0})}\right]
\end{aligned}
\tag{11.59}
$$

对公式（11.59）计算出的运行状态的概率求和，可以得出与公式（11.57）相同的可靠度表达式。

11.5 容错计算机产品的覆盖建模

如今，微处理器产品得到广泛的应用，从汽车到厨房用品再到缝纫机都可以看到微

处理器的身影。此类应用中的计算机失效非常麻烦，但还不至于威胁生命。计算机产品也经常应用于关键性领域，例如飞行器控制、核电厂监测、化学处理或者医疗领域等。因为在这些产品关键部分的失效可能使操作者丧失生命或者将产生严重的环境问题和经济损失，因此这类产品需要有非常高的可靠性。本节将主要讨论容错计算机（Fault-Tolerant Computer，FTC）产品的可靠性问题。

容错计算机产品经常被设计成包含足够数量的处理器，以减小所有处理器同时失效的概率。技术的进步已经使这类现象有所减少。但是，产品设计师仍然必须确保使用者能迅速发现失效和错误，以便可以更有效地发挥冗余单元的作用。如果失效单元不能够在产品的重组中排除出来，那么它将产生异常结果，并损害其他非失效单元。容错计算机产品的可靠性模型包括有覆盖因素（Coverage Factor），以反映产品在运行过程中从发生失效到自动恢复的能力。覆盖因素表示产品从故障及其相关错误中自动恢复，并继续正常运行的概率。

即使备用单元仍然正常，FTC 产品也可能无法从故障中恢复。例如故障可能会产生一个未发现错误，随后的计算或操作将在错误的数据上进行，并最终导致整个产品失效。即使错误被发现，产品也可能会因为此错误已经损坏自动修复程序而不能使失效部件自动恢复。即使备用或冗余单元仍然正常，产品不能自动从中恢复的故障——即未恢复故障（Unrecovered Fault）——也将使产品立即失效。

11.5.1 相关术语

已覆盖故障（Covered Fault）——来自于可以自动恢复产品的故障。虽然产品可能处于功能退化模式，但是产品仍可以继续运行。

错误（Error）——计算产品处理信息过程中的故障表现或者其内置产品状态的故障表现。

失效（Failure）——预期服务的一种不可接受偏差，一个异常的输出，一种不能实现期望功能的无能（Inability）。

故障（Fault）——某些硬件或软件中的缺陷、不完美因素或者瑕疵。

FTC——容错计算机。

硬件故障（Hardware Fault）——在硬件生命周期内出现的故障。例如两导线之间的短路或晶体管连接中的断路。

永久性硬件故障（Permanent Hardware Fault）——具有持续影响的物理故障。

瞬时硬件故障（Transient Hardware Fault）——引起硬件非永久性损伤的有时限的效物理故障。瞬时故障一般由过热、动力中断或者环境影响产生。

未覆盖故障（Uncovered Fault）——产品不能从中自动恢复的故障。未覆盖故障将导致产品立即失效。

11.5.2 不完全覆盖的影响

案例 11.1 故障覆盖的建模

某 FTC 产品有四个处理器。只要不出现未覆盖故障，产品就可以持续运行，直到最后一个处理器失效时才失效。但是，任何处理器中的未覆盖故障都能使产品立即失

效。图11.30是此示例产品的Markov模型。其中，λ表示单个单元的失效率，c_p表示覆盖率。

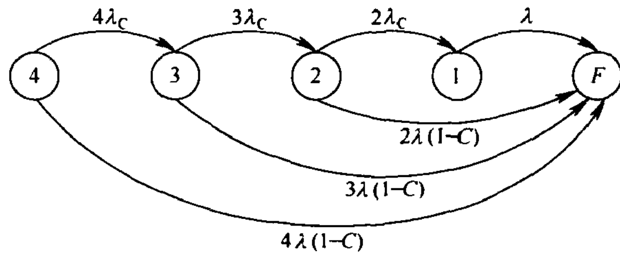


图11.30　四冗余单元系统所覆盖故障的Markov模型

当失效率为$\lambda=10^{-4}$每小时，覆盖率为$c_p=0.99$时，产品在100运行小时内的失效概率为3.98×10^{-4}。对于同一个产品，假如覆盖了所有的故障（$c_p=1$），那么产品失效的概率将相当低，为9.8×10^{-9}。未覆盖故障对产品产生巨大影响的概率只有百分之一。假如产品有三个单元而不是四个，那么失效的概率（此时$c_p=0.99$）将是2.98×10^{-4}，这比四个冗余单元的情况还要好。

11.5.3　覆盖模型的一般结构

案例11.1中的简单的四冗余单元产品表明：不完全覆盖能对容错产品的可靠性产生巨大影响。对于如何合理地确定产品模型中的c_p值，相关研究人员已经采用过多种方法。假如有产品的工作模型或原型或足够的产品设计信息，那么我们就可以构建覆盖模型。模型的参数可以从工作原型测量得到，也可以从使用现场数据中估计，或者直接构建出产品恢复过程的详细模型。假如恢复过程的细节未知，那么我们可以从其他类似产品中推测出合理的参数。本节将着重讨论故障出现时产品的详细行为的建模。

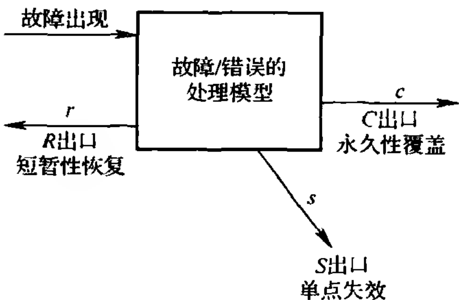


图11.31　覆盖模型的一般结构

如图11.31所示为覆盖模型的一般结构。模型的进入点表示故障的出现，三个出口（R，C和S）是三个可能的输出结果。

R：瞬时恢复（Transient Restoration）。正确识别瞬时故障，并从中恢复。瞬时故障通常由外部因素或环境因素引起，例如过热或电源电路上的假信号。大多数故障都是瞬时的。从瞬时故障中成功恢复，可以使产品复原到运行状态而不丢弃任何部件，例如通过掩蔽错误、重试指令或者重新回到前一个监测位置来恢复产品。为了成功地到达出口，需要完成以下工作：

① 及时检测故障产生的错误。

② 有效恢复程序的表现。

③ 快速消除故障（错误产生的原因）。

C：永久性覆盖。决定失效的永久性本质、成功地隔离和移除有故障的部件。

S：单点失效。通常，当未发现错误在产品中蔓延或者当有故障单元不能被隔离，并且产品不能重组的时候，由单个故障引发的产品失效。

案例11.2　内存储器子产品的覆盖模型

内存储器子产品的一个假设恢复过程如图11.32所示。因为内存储器使用了错误校

正码（Error-Correcting Code），所以单个位错误（Single-Bit Error）经常能够被检测到，并能得以纠正，产品不需要重组。假如所有98%的内存储器故障仅影响一个单位（Single-Bit），那么到达 R 出口的概率是 $r=0.98$；影响更多存储位的2%的内存储器的故障被发现的概率是95%。当检测到一个多内存错误（Multiple Memory Error）时，系统将抛弃受影响的内存，然后更新内存的映射功能，并从前一个检测点处重新加载所需信息，更新当前的产品状态。在原型产品上的试验发现，这种从检测到的多内存错误中恢复系统成功的概率只有85%。因此，到达 C 出口的概率就是发现多重故障出现并从中恢复的概率，即

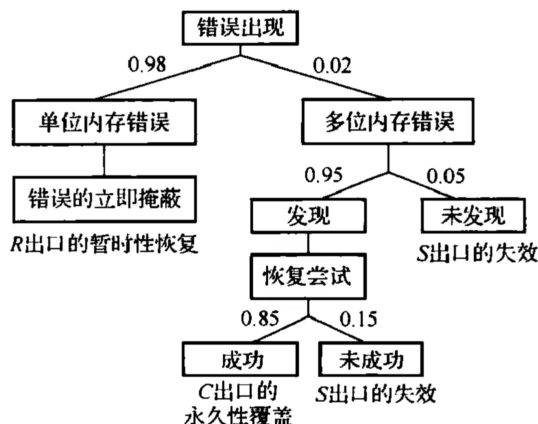


图 11.32　内存子系统的覆盖模型

$$c_p=(0.02)\times(0.95)\times(0.85)=0.11615$$

通向单点失效出口的两条路径是：

① 假如没有发现多位错误（Multiple-Bit Error）（其概率为 0.02×0.05），那么内存故障将引起单点失效。

② 虽然发现了多位内存错误，但尝试性恢复不成功（其概率为 $0.02\times0.95\times0.15$）。

因此，$s=(0.02)\times[(0.05)+(0.95)\times(0.15)]=0.00385$。

案例 11.3　处理器的覆盖模型

处理器包含内置测试电路，它的作用是在执行指令的同时检测错误的发生。假如监测到了错误，那么立刻重新尝试执行指令。为防止指令重试不成功，将部分执行结果存储起来，因此，计算仍然可以从进程中的一些中间点（检测位置）继续执行，此过程称为返回重新执行（Rollback）。在某些情况下，故障过于严重会导致返回重新执行不成功，此时，计算过程必须在启动系统级恢复程序后重新执行。处理器的故障覆盖模型和后续恢复程序如图 11.33 所示［Ng and Avizienis，1976］。

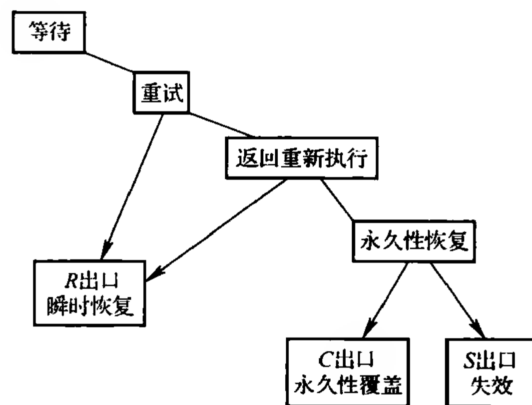


图 11.33　处理器的覆盖模型

① 瞬时恢复程序（Transient Recovery Procedure）。假设故障是瞬时的，只要发现错误就开始执行一个多步骤恢复程序。如果在执行了三个步骤之后故障仍然存在，那么必须启动永久性恢复程序。

步骤 1：等待 0.1s 不做任何事情。假如故障是瞬时的，那么它将在此段时间内消

失，并使得系统可以返回重新执行。

步骤2：多次重新尝试当前的指令约0.5s。重试成功（即没有检测到错误）的概率为0.5。

步骤3：假如错误持续，在2s内返回到前一个检测点重新执行计算。返回重新执行成功并清除错误的概率为80%。

假如故障是瞬时的，那么只要故障在以上步骤开始执行前消失，瞬时恢复将会成功。为了分析这个例子，假设瞬时故障的寿命服从指数分布，均值为0.25s。90%的故障都是瞬时的，而另外10%是永久性的。

② 永久性恢复（Permanent Recovery）。假如错误在返回重新执行后仍然持续存在，则它可能是由永久性故障引起的。系统级永久性故障恢复过程将开始从工作单元集中移除已损坏的处理器，并在此之后重组。永久性故障恢复过程成功的概率为0.875。

分析此覆盖模型，包括计算三步瞬时恢复的每一步产品恢复（PSR_i）概率和永久恢复概率。此计算需要决定两个中间级的数量集合：到达步骤i之前瞬变的概率和步骤i发生的概率。

③ 瞬时恢复出口。假如失效是瞬时的，且三个步骤中的任何一个使产品成功恢复，则系统可以到达瞬时恢复出口。

步骤1：该步骤立即执行的概率取决于故障的出现情况。该步骤并没有执行实质性的恢复，所以产品在该步骤后成功恢复的概率是零（$PSR_1=0$）。

步骤2：如果瞬时故障已在步骤1中消失（概率是$1-e^{-0.1/0.25}=0.329$），并且重新执行指令成功（其概率是0.5），产品将在该步骤中从瞬时错误中恢复。

因此，$PSR_2=(0.329)\times(0.5)=0.165$。

步骤3：如果步骤1与步骤2都不成功（其概率是$1-0.165=0.835$）、瞬时故障已在步骤1和步骤2中消失（概率是$1-e^{(-0.1+0.5)/0.25}=0.909$）、返回重新执行成功（其概率为0.8），产品将在该步骤中从瞬时故障中恢复。

因此，$PSR_3=(0.835)\times(0.909)\times(0.8)=0.607$。

产品从瞬时故障中恢复的概率等于以上三个步骤的概率之和（$0+0.165+0.607=0.772$）。假如故障是瞬时的（概率是0.9），并且瞬时恢复成功，系统就可到达瞬时恢复出口。因此，$r=0.9\times0.772=0.695$。

④ 永久性恢复出口。在分析永久性恢复出口时要考虑两个事件：第一，启动永久性恢复程序来处理持续瞬时故障；第二，从永久性故障中恢复系统。

事件1：假如瞬时恢复的三个步骤都不成功，那么将启动永久性恢复程序来处理瞬时故障。与此事件关联的概率是以下各项概率的乘积：故障是瞬时性的概率（0.9）、瞬时恢复程序的三个步骤均不成功的概率（$1-0.772=0.228$）、永久性恢复过程成功的概率（0.875）。

事件2：如果故障是永久性的（概率为0.10）、永久性恢复过程成功（概率是0.875），那么永久性恢复过程就可以成功处理永久性故障。

系统到达永久性恢复出口的概率是此两种情况的概率之和，即$c_p=0.179+$

0.875 = 0.267。

⑤ 单点失效出口：分析单点失效出口需要考虑两个事件。

事件 1：对于瞬时故障，假如永久性恢复程序启动，但是没有使系统恢复，那么系统就会到达单点失效出口。与此情况相关的概率：0.228 × (1 − 0.875) = 0.028。将其与故障属于瞬时故障的概率相乘，将得出事件 1 的概率：0.9 × 0.028 = 0.0252。

事件 2：对于永久性失效，假如永久性恢复程序不成功，那么系统就会到达单点失效出口。将永久性恢复程序不成功的概率与故障属于永久性故障的概率相乘，将得出事件 2 的概率：(1 − 0.875) × (0.10) = 0.0125。

系统到达单点失效出口的概率是事件 1 和事件 2 的概率之和：$s = (0.0252 + 0.0125) = 0.0377$。

11.5.4 近重合故障

在高可靠度产品（例如用于控制飞行的产品）中，当产品从前一个故障中恢复后，第二个故障出现的概率将可能决定产品的失效概率。第二个近重合故障（Near-Coincident Fault）将可能干扰恢复过程，并使产品立即失效。

案例 11.4　完善恢复过程的指数分布时间—非瞬时故障

此案例利用一个非常简单的覆盖模型来说明近重合故障所产生的影响。恢复过程假设每个故障都是永久性的，并且恢复可以重组产品以绕开有故障的部件。此种情况下，不尝试执行瞬时恢复程序。只要没有第二个故障的干扰，恢复常常都会成功，系统将到达永久性恢复出口。恢复时间呈指数分布，速率为 δ_r，在第二个故障出现所用的时间也呈指数分布，速率为 γ。假如恢复时间少于直到下一个故障出现所用的时间，那么将覆盖第一个故障（概率为 c_p）。

$$c_p = Prob[\text{系统恢复所用时间} < \text{第二个故障出现所用时间}]$$

$$= \int_0^{\infty} (\delta_r e^{-\delta x})(e^{-\gamma x})\,dx + \frac{\delta_r}{\delta_r + \gamma} \qquad (11.60)$$

一般情况下，系统从故障中成功恢复有两个条件：

① 在时间充足的情况下，产品能够从故障中恢复。

② 必须在下一个故障的干扰之前执行恢复程序。

使用第 11.1.2 节中所介绍方法可以分析第一个条件。要计算干扰故障出现的概率，就需要了解产品需要多长才能时间恢复。

为了确定近重合故障出现的概率，我们需要一些新的计算符号，因为到达出口所需的时间必须包括在到达出口的概率中：

$P_C(\tau)$ ——系统在小于等于故障出现所需的时间 τ 内，通过少数部件恢复到某个状态（即到达覆盖模型中的 C 出口）的概率；

$P_R(\tau)$ ——在小于等于故障出现所需时间 τ 内，瞬时修复成功的概率；

$P_S(\tau)$ ——在小于等于故障出现所需时间 τ 内，系统出现单点失效的概率。

三种分布 $P_C(\tau)$、$P_R(\tau)$ 和 $P_S(\tau)$ 可能会存在缺陷，因为它们的极限值可能小于 1。实际上，

$$\lim_{\tau \to \infty}[P_C(\tau) + P_R(\tau) + P_S(\tau)] = 1 \tag{11.61}$$

这种分布是不考虑近重合故障时的覆盖模型的解。

设 $\hat{r}$、$\hat{c}$ 和 $\hat{s}$ 表示到达各个出口的极限概率：

$$\hat{r} = \lim_{\tau \to \infty} P_R(\tau), \hat{c} = \lim_{\tau \to \infty} P_C(\tau), \hat{s} = \lim_{\tau \to \infty} P_S(\tau), \hat{r} + \hat{c} + \hat{s} = 1 \tag{11.62}$$

如果忽略近重合故障出现的概率（如同第11.1.2节的假设），那么，期望的恢复因子（r、c 和 s）等同于这些极限概率：

$$r = \hat{r}, c = \hat{c}, s = \hat{s} \tag{11.63}$$

在某些产品中，仅了解产品的最终恢复状态是不够的。只有在第二个故障（即干扰故障）出现之前完成恢复过程，恢复才能算是成功的。因此，用覆盖因子来作为出口分布的极限值是比较可行的，特别是当恢复时间比故障出现所用时间长的情况下，这种做法尤为有效。在考虑系统从故障中恢复所需的时间时，需要对这些极限值加以调整。

设 W 是表示干扰故障出现的间隔时间的随机变量，$F_W(t) = 1 - e^{-\lambda t}$。设 Y_R、Y_C 和 Y_S 是表示系统到达相应出口所需时间的随机变量，$F_{YR}(\tau)$、$F_{YC}(\tau)$ 和 $F_{YS}(\tau)$ 到达出口所需时间的条件分布：

$$F_{YR}(\tau) = \frac{P_R(\tau)}{\hat{r}}, F_{YC}(\tau) = \frac{P_C(\tau)}{\hat{c}}, F_{YS}(\tau) = \frac{P_S(\tau)}{\hat{s}} \tag{11.64}$$

这些分布是无缺陷的分布，因为每个极限值都等于1。与之对应的概率密度函数是 $f_{YR}(\tau)$、$f_{YC}(\tau)$ 和 $f_{YS}(\tau)$。

P_R 和 F_R 分布是不同的，$P_R(\tau)$ 是从模型入口到达 R 出口所需时间的分布。在极限情况下，P_R 的值可能不等于1，因为 R 出口可能不会出现，而系统会到达 C 和 S 出口。另一方面，F_R 分布是系统真正能到达 R 出口所需时间的分布。在考虑完整的模型时，P_R 分布才能真正表示到达 R 出口所需时间的分布，而 F_R 仅考虑能到达 R 出口的部分，忽略了模型的其余部分。假如把另一个出口加入到模型中，那么 P_R 分布将有所改变，而 F_R 分布则不会。

覆盖因子的计算必须包括近重合故障出现的概率。为了分析永久性恢复出口 C，系统必须到达 C 出口（而不是另两个出口），且必须在另一个故障出现之前到达。系统到达 C 出口的概率 c_p 由式（11.65）计算：

$$\begin{aligned} c_p &= Prob[\text{系统到达 } C \text{ 出口,且所需时间} < W] \\ &= \int_0^\infty (e^{-\gamma x}) \mathrm{d}P_c(x) = \hat{c} \lim_{u \to \gamma} L_{YC}(u) \end{aligned} \tag{11.65}$$

其中，$L_{YC}(u)$ 是随意变量 Y_C 的 Laplace 变换。使用同样的方法可以计算出 r 和 s。

在第11.1.2节的计算中，$\hat{r} + \hat{c} + \hat{s} = 1$，它忽略了近重合故障。当考虑近重合故障时，就需要为覆盖模型再加入第四个出口（N）。N 出口的覆盖因子是 $n = 1 - (r + c + s)$。

因为近重合故障出现的速率取决于系统的完整配置（产品的部件越多，近重合故障出现的概率就越高），覆盖参数是近重合故障出现速率 n 的函数。n 的数值取决于完整的产品模型（见第11.1.4节）。

让我们回到前面的覆盖模型。对于内存储器的覆盖模型（案例 11.2、图 11.32），其覆盖因子为

$$\hat{r}=0.98,\hat{c}=0.01615,\hat{s}=0.00385 \tag{11.66}$$

瞬时恢复出口：因为瞬时恢复几乎是瞬间完成的，因而此过程中不可能有近重合故障出现，则

$$r(\gamma)=\hat{r}=0.98 \tag{11.67}$$

永久性恢复出口：永久性恢复所需的时间服从参数为 δ_r 的指数分布，所以

$$c(\gamma)=0.01615\times\frac{\delta_r}{\delta_r+\gamma} \tag{11.68}$$

单点失效出口：因为单点失效是由永久性恢复失效引起的，所以单点失效时间的分布与永久性失效一样。因此

$$s(\gamma)=0.00385\times\frac{\delta_r}{\delta_r+\gamma} \tag{11.69}$$

近重合失效出口：近重合失效出口的覆盖因子是

$$n(\gamma)=1-[r+c(\gamma)+s(\gamma)] \tag{11.70}$$

对于处理器的覆盖模型（例 11.3、图 11.33），覆盖因子的计算稍有差异。成功恢复系统的概率由系统进入相位的概率、瞬时故障消失的概率以及相位的有效性三者的积确定，但现在还要乘以近重合故障不在相位内出现的概率。

11.5.5　把覆盖模型纳入到产品模型

一旦确定了产品模型和覆盖模型，就可以使用行为分解（Behavioral Decomposition）将覆盖模型的结果融入到产品模型中。本节介绍将覆盖模型的结果融入产品失效模式的 Markov 模型中的方法。

案例 11.5　三个处理器，两个内存的 FTC 产品

某简单 FTC 产品（称其为 3P2M）由三个处理器和两个通过共享总线连接的内存组成（见图 11.34）。只要有一个处理器能够与一个内存连接，那么产品就可以运行。图 11.35显示了 3P2M 产品的 Markov 链，其状态记为一个有序三元标记：

① 处于运行状态的处理器数量。

② 处于运行状态的内存的数量。

③ 总线的状态。

部件的失效率常量是：

① 处理器：$\lambda=10^{-4}$。

② 内存：$\lambda=10^{-5}$。

③ 总线：$\lambda=10^{-6}$。

在图 11.35 所示的 Markov 链中，

$F1$——处理器群衰竭（Exhaustion）；

$F2$——内存衰竭；

$F3$——总线失效。

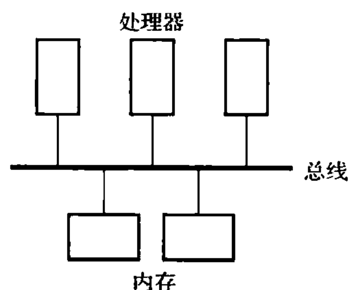


图 11.34　简化后的组合系统及其覆盖模型

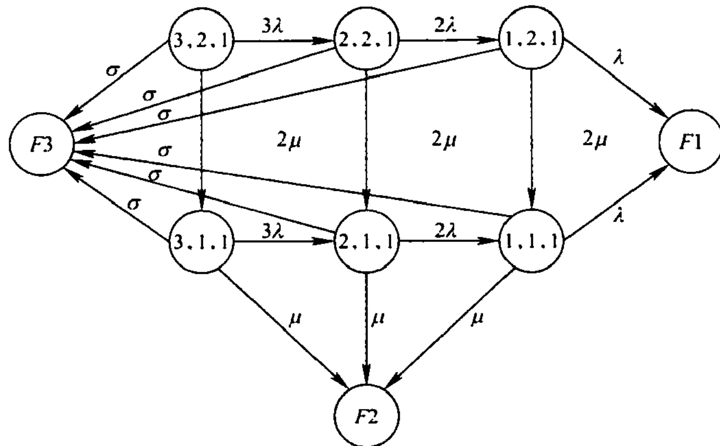


图 11.35　3P2M 系统框图

处理器的覆盖模型与案例 11.3（见图 11.33）的相同，内存的覆盖模型与案例 11.2（见图 11.32）的相同，总线故障不能自动恢复。覆盖模型被插入到 Markov 链中运行状态之间的每个圆弧上，如图 11.36 所示。在 3P2M 产品中，处于水平圆弧上的覆盖模型是相同的处理器覆盖模型（见图 11.33），处于垂直圆弧上的覆盖模型是相同的内存覆盖模型（见图 11.32）。插入的两个额外失效状态是：

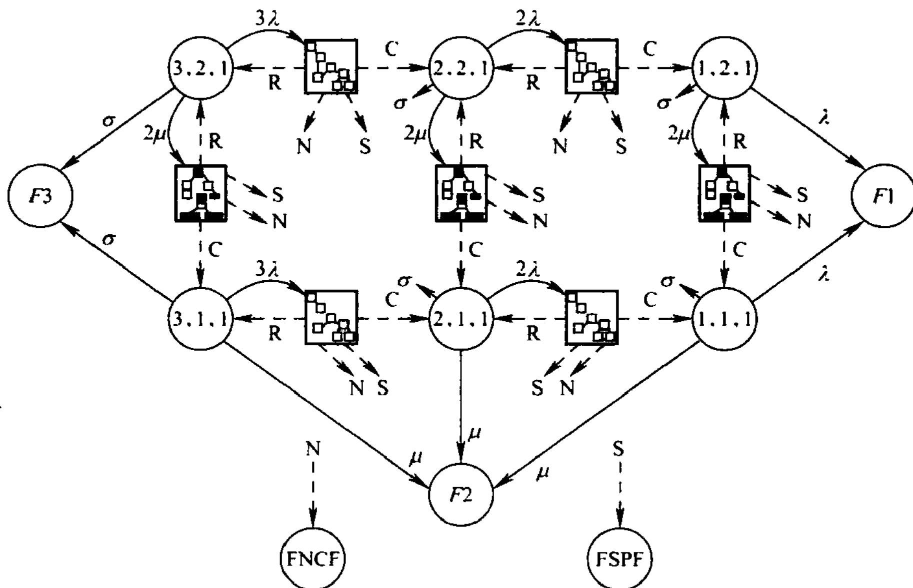


图 11.36　3P2M 案例系统的 Markov 链模型

FSPF——出现单点失效；

FNCF——出现几乎成对的近重合失效。

解覆盖模型的解是系统到达每一个出口的概率。用覆盖参数代替插入到圆弧上的覆盖模型，所得的 Markov 链（见图 11.37）就是产品的状态概率。产品可靠性就是产品

不处于任何失效状态的概率。

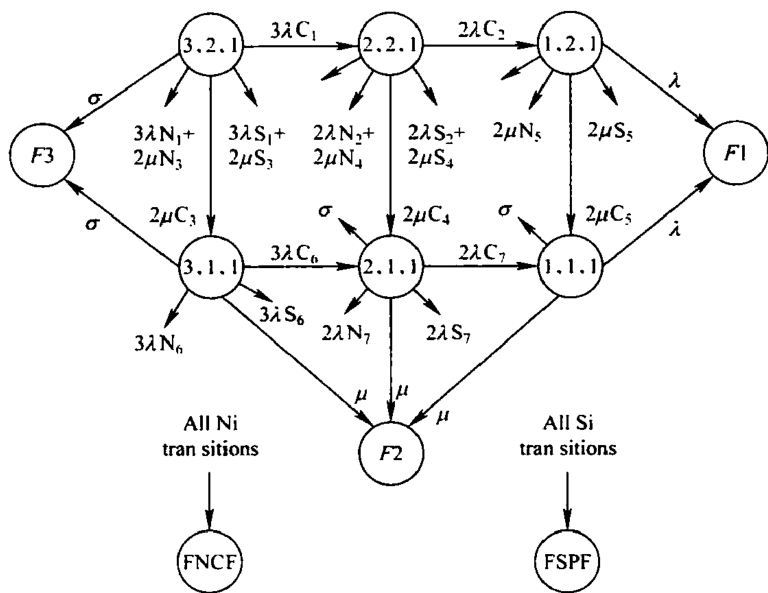


图 11.37　简化后的组合系统及其覆盖模型

因为每个状态的工作部件数量都不同，所以近重合失效出现的速率与产品所处状态和状态的转换相关，这也导致每次状态转换时的覆盖因子不同。例如圆弧上起始于状态“3，2，1”的两个覆盖模型。在状态“3，2，1”中，一个处理器出现故障，其他两个处理器、两个内存和总线依然处于工作状态。在这些工作部件中的任何故障都会干扰恢复过程，因而与处理器恢复模型相关的近重合故障率是

$$\gamma_1 = 2\lambda + 2\mu + \sigma \tag{11.71}$$

从相同状态开始，当内存失效发生时，三个处理器、一个内存和总线仍旧是起作用的，因此与内存覆盖模型相关的近重合故障率是

$$\gamma_2 = 3\lambda + \mu + \sigma \tag{11.72}$$

表 11.2 列出了 3P2M 案例的 100h 运行的可靠度分析结果。单点失效是不可靠度的最大因素，单点失效中的故障是不可恢复的。

表 11.2　3P2M 案例产品的解

失效原因	概　率
处理器衰竭	2.00×10^{-7}
内存衰竭	1.61×10^{-8}
总线衰竭	9.99×10^{-5}
单点失效	3.53×10^{-4}
近重合失效	4.23×10^{-9}
总不可靠度	4.53×10^{-4}

11.6 有界近似模型

描述为故障树的产品的失效（顶事件发生的概率）概率的计算量非常大。Markov模型将随着产品的构成部件的数量呈指数性增长。我们只能精确分析那些拥有相对较少部件，且几乎没有或者只有很少的子产品是相互关联的产品的可靠度。

在本节中，我们将介绍一些求解容错产品模型近似解的方法。这些方法缩短了求解过程，近似解表示为产品可靠度的上边界和下边界，所以我们可以估算近似解的近似程度。此处介绍两种简化故障树的组合解的方法。当假设覆盖模型是完整的，且可以忽略近重合故障时，这两种方法非常有效。对于那些能够用 Markov 模型精确建模的产品，必须要考虑简化其 Markov 模型。我们还将介绍用于分析短期运行且具有高可靠度产品的方法，这类产品的短期运行中，不能对产品进行手动维修。这种短期运行导致产品同时发生多个失效的概率非常小。组成产品的部件通常都有恒定的失效率，但此处将介绍用于分析其失效过程与时间相独立的产品的技术。

案例 11.6　*Cm 产品**

*Cm** 产品是一种松弛耦合分布（Loosely Coupled Distributed）产品，此处用它来描述前两种简化方法。图 11.38 所示的 *Cm** 产品案例包括两个簇（Cluster），每个簇包含四个处理器和四个内存模块。假定故障恢复及时，且效果良好，因此，此处不需要使用覆盖模型。故障树模型比 Markov 链更适合用来分析此类产品。

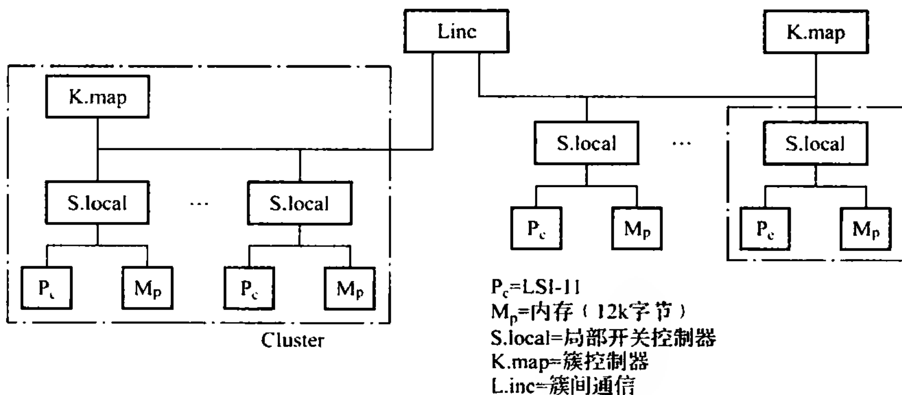


图 11.38　*Cm** 产品的框图

每一个簇包括四个局部开关控制器（S. local），并且每个控制器都连接到一个处理器和一个 12k 的内存模块。每一个处理器在主板上有 4k 内存。K. map 是一个与 S. local 连接的簇控制器，簇与簇通过簇间通信（L. inc）连接起来。K. map 中的故障会使它与 S. local（以及处理器和内存）的连接断开，而 S. local 中的故障会使与它与处理器和内存模块的连接断开。

失效率常数为：

内存：8 个内存模块，每个模块的失效率为 $\lambda_M = 69.4$ 每百万小时；

处理器：16 个处理器，每个处理器的失效率为 $\lambda_P = 29.9$ 每百万小时；

S. local：8 个局部开关控制器，每个控制器的失效率为 $\lambda_S = 24$ 每百万小时；

K. map：两个簇控制器，每个控制器的恒定失效率为 $\lambda_K = 131$ 每百万小时；

L. inc：一个簇间通信连接，其恒定失效率为 $\lambda_L = 34.8$ 每百万小时。

产品需要三个内存与三个处理器通信才能正常工作。只要 L. inc 可运行，那么两个簇的部件可以满足产品的需求。但是，假如 L. inc 失效，那么一个簇就必须满足产品的需求。*Cm** 产品的故障树模型如图 11.39 所示。产品失效（顶事件）有一半归因于通向最高的 OR 门的输入。当 L. inc 失效，且单个簇无法满足产品需求时或者 L. inc 的状态是独立的，但两个簇的处理器或内存都不足以满足产品需求时，失效就会发生。当 n 个输入中有 m 个都已出现时，标签为 m/n 的逻辑门就为真，m 和 n 都是整数。故障树底部的一排 OR 门反映了以下情况：当一个处理器（或一个内存）已失效或者 S. local 失效或 K. map 失效时，处理器（或内存）就会失效。

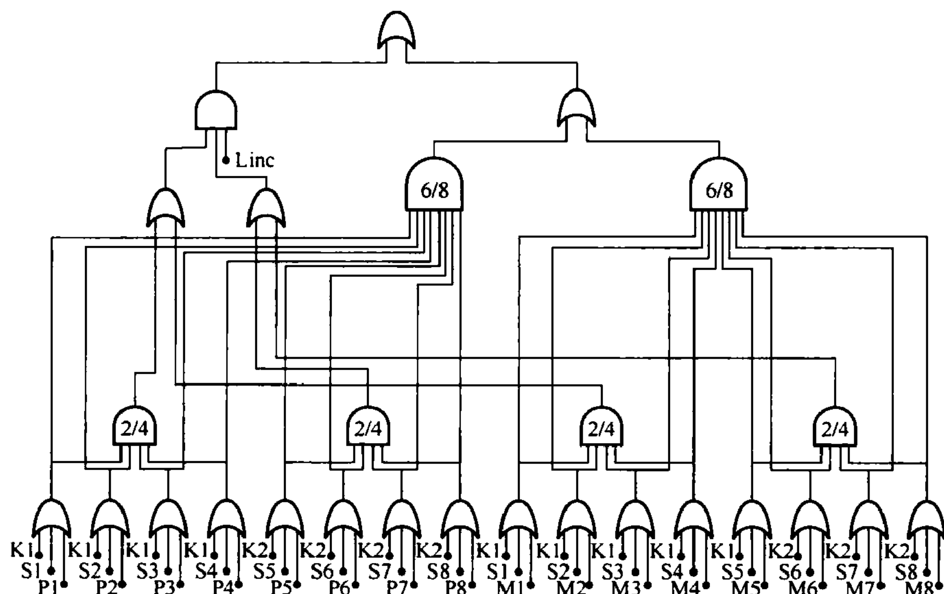


图 11.39　*Cm** 系统的故障树模型

此案例产品有许多可能的状态。因为它有 27 个部件，假如任何部件处于工作或失效状态，那么产品将可能处于 $2^{27} > 13.4$ 亿个状态的任何一个，此产品有 5405 个最小割集。

11.6.1　截断穷尽状态枚举

对于已知的产品故障树模型，最简单的求解方法是穷尽状态枚举（Exhaustive State Enumeration）。使用这种方法时，我们将检测并确定所有 2^n 个状态矢量（假如有 n 个基本事件），以表示产品运行或者失效时的配置。对与失效配置的状态矢量相关的概率求和，就可以得到产品的不可靠度。虽然这样做不是很有效，但是穷尽状态矢量检测过程可以很好地简化过程本身。在求解过程的任何位置，我们可以界定与未检测状态矢量有关的总概率。当这个总额小于所要求的精度时，状态枚举过程可以停止，同时也就得出

了可靠度的有界估计值。

假设在 t 时刻，n 个基本事件以失效概率的降序排列，用以求解产品的预期可靠度。也就是说，部件 i 的失效概率高于部件 $i+1$ 的失效概率，其中，$1<i<n$。长度 n 的二进制矢量用来表示部件的状态。如果某部件处于状态矢量 i，那么就意味着部件 i 已经失效；0 表示部件 i 是可运行的。产品的初始状态（0，0，0，…，0）表示所有部件都没有失效。从初始状态开始直到（1，1，1，…，1）矢量，按照排序惯例对状态矢量进行有序评估。每个状态矢量均表示产品的可运行配置或失效配置（与产品故障树模型所定义的一样）。所有可运行配置出现的概率之和是产品的可靠度，所有失效配置出现的概率之和就是产品的不可靠度。

每一个状态矢量出现的概率由单个部件的失效概率决定。用 p_i 表示第 i 个部件的失效概率，q_i 表示第 i 个部件没有失效的概率。因为两个事件是互补的，所以 $p_i+q_i=1$。设 v_i 表示正在分析的状态矢量的第 i 个元素（每个 ν_i 是 0 还是 1 取决于第 i 个部件是运行的还是失效的）。状态矢量（ν_1，ν_2，…，ν_n）出现的概率由式（11.73）给出

$$Prob[(\nu_1,\nu_2,\cdots,\nu_n)] = \prod_{i=1}^{n}[\nu_i p_i + (1-\nu_i)q_i] \tag{11.73}$$

此节的目的是描述一个过程，通过此过程，我们可以不必分析所有 2^n 种产品状态，却可以估计出具有某种精确度（例如 10^{-d}）的产品可靠度。某些关键状态矢量出现的概率可以决定为达到 10^{-d} 精度而必须检测的失效级别数。因为每个失效级别的错误经常被高估，所以估计结果的精度往往比所要求的精度高。失效级别表示所有在失效级别 k 上的状态矢量都有一个精确的 k 值和 $n-k$ 值，这些值表示 k 个部件已经失效，而其他部件仍然运行。

因为基本事件以失效概率的降序排列，所以任何失效级别上的最可能状态矢量都是（1，…，1，0，…，0）。将级别（n，1）上状态矢量的数目与在级别 l 上的第一个状态的相关概率相乘，就可得到此级别上所有状态失效概率的最大可能值。一般来说，在任何失效级别 l 上，由于失效级别上未扩展状态导致的最大错误是：

$$E_1 = \binom{n}{l} p_1 p_2 \cdots p_i q_{i+1} \cdots q_n \tag{11.74}$$

为了避免检测所有 2^n 个状态矢量，但同时要保证 10^{-d} 的精度，$E_i s$ 之和必须小于 $5\times10^{-(d+1)}$。对于每个用于计算总和的 E_i 来说，当它生成状态矢量时，我们就可以忽略相应的失效级别 i。一个实用的解决方案是找出最小的 l，以使式（11.75）成立：

$$Error\ Term = \left(\sum_{i=l+1}^{n} E_i\right) \leqslant 5\times10^{-(d+1)} \tag{11.75}$$

满足公式（11.75）的最小 l 值决定了需要在多少个级别（l）上展开产品状态，以达到所需要的精度。对于一个已知满足公式（11.75）的 l 值，我们必须分析所有与包含 l 或者更少失效部件数的状态相对应的状态矢量，因此，与包含 $l+1$ 个或更多失效部件的状态相对应的状态矢量不需要展开。因为与每个失效级别相关的错误都被高估了，所以所有错误发生的概率之和将超过 1。

不可靠度的界限可以由式（11.76）描述。设 U_i 是级别 i 上所有失效配置的状态矢量概率之和，则产品精确的不可靠度是

$$Unreliability \sum_{i=1}^{n} U_i \tag{11.76}$$

其边界是

$$\sum_{i=0}^{n} U_i \leqslant Unreliability \leqslant \sum_{i=0}^{n} U_i + \sum_{i=l+1}^{n} E_i \tag{11.77}$$

表 11.3 列出了通过使用截断穷尽状态枚举算法（Truncated Exhaustive State Enumeration Algorithm）得出的 Cm^* 产品的所有解。合理精确的估计产品可靠度源于对一小部分状态空间的考虑。随着运行时间的增加，越来越多的部件可能会失效，所以，我们应该考虑更多级别的状态空间。

表 11.3　Cm^* 产品的故障树的解

简化穷尽状态枚举法 （Cm^* 产品的故障树——总共有 13 421 778 个状态）				
运行时间/h	5	10	100	1000
简化级别（部件失效数量）	3	3	4	7
模型大小（状态数量）	3303	3303	20853	1285623
不可靠度下边界	$4.30e^{-7}$	$1.73e^{-5}$	$1.85e^{-4}$	$2.65e^{-2}$
不可靠度上边界	$4.30e^{-7}$	$1.75e^{-5}$	$1.90e^{-4}$	$2.84e^{-2}$
总运行时间 （Sun-4 型 CPU，s）	28	27	168	8923

11.6.2　截断的不相交积之和

大部分用于确定建模故障树的产品的不可靠度的定性算法都从确定产品的最小割集开始。一个割集是一组基本事件。在这个事件组中，假如所有基本事件发生，那么顶事件（产品失效）将会发生。假如任何基本事件从最小割集中移除，那么剩余的事件就不再是一个割集。假如最小割集 p 的标签为 C_i（$i=1,\cdots,p$），那么产品的不可靠度是：

$$Prob\left[\bigcup_{i=1}^{p} C_i\right] \tag{11.78}$$

割集不一定都是互不相交，所以我们不能简单地将独立割集的概率相加，但是相加后的和却能确定产品不可靠度的上边界。

在用于故障树评价的不相交及之和（Sum of Disjoint Products，SDP）所使用的公式中，$\overline{C}_i$ 是 C_i 之外的通用集（Universal Set）。因为公式（11.79）右侧项的互不相交，所以对独立项概率求和，就可以精确求出产品的不可靠度。

$$\bigcup_{i=1}^{p} C_i = (C_1) \bigcup (\overline{C}_1 C_2) \bigcup (\overline{C}_1 \overline{C}_2 C_3) \bigcup \cdots \bigcup (\overline{C}_1 \overline{C}_2 \overline{C}_3 \cdots \overline{C_{p-1}}\, C_p) \tag{11.79}$$

一般来说，除了拥有最小配置的产品之外，我们无法确定其他产品的最小割集，也

无法用最小割集来计算其他产品的可靠度。通常，计算割集要比计算可靠度快，因此，本小节主要讨论如何在确定割集之后截断 SDP 的问题。

SDP 算法的基本方法是使用布尔代数（Boolean Algebra）表达每个割集，并使其与前一个割集不相交。假如割集以出现概率降序排列，则产品不可靠度的最大影响因素是那些拥有较低指数的割集。使与那些拥有最高指数的割集与它前面的割集互不相交，不仅不能影响可靠度的精度，还将耗费大量的时间。在 Cm^* 产品的案例中，最可取的割集（同时包含两个 Kmap）是按照重要程度排序的第四个割集。这些在割集之间的差异可以使我们不用去让其他不可取的割集与其他割集不相交，但却能估计出产品不可靠度的边界。

SDP 算法也可以自行截断。假设前 l 个割集（$l<p$）已经与它们前面的割集不相交，即

$$\overline{C}_1C_2,\overline{C}_1\overline{C}_2C_3,\cdots,\overline{C}_1\overline{C}_2\overline{C}_3\cdots\overline{C_{l-1}}\,C_l \tag{11.80}$$

所有的割集都已确定，那么这些割集将与剩余的割集（从 $l+1$ 开始到 p）用来确定产品的不可靠度：

$$\begin{aligned}&Prob[C_1]+Prob[\overline{C}_1C_2]+Prob[\overline{C}_1\overline{C}_2C_3]+\cdots+Prob[\overline{C}_1\overline{C}_2\cdots\overline{C_{l-1}}C_l]\\&\leqslant Unreliability\leqslant\\&Prob[C_1]+Prob[\overline{C}_1C_2]+Prob[\overline{C}_1\overline{C}_2C_3]+\cdots+Prob[\overline{C}_1\overline{C}_2\cdots\overline{C_{l-1}}C_l]\\&+Prob[C_{l+1}]+Prob[C_{l+2}]+\cdots+Prob[C_p]\end{aligned} \tag{11.81}$$

假如在使割集 l 不相交以后，边界已经足够严格，那么终止使剩余割集不相交的过程。假如从 l 个相交割集得来的边界太松弛，那么继续让第 $l+1$ 个割集与它前面的割集不相交。实际上，在误差区间的最大宽度已知的情况下（例如 10^{-d}），我们就可以确定 l 前面的割集。和在第 11.6.1 节中一样，我们取割集的概率之和小于 $5\times10^{-(d+1)}$。

例如求解最小的 l，使式（11.82）成立

$$\left(\sum_{i=l+1}^{p}Prob[C_i]\right)\leqslant 5\times10^{-(d+10)} \tag{11.82}$$

满足公式（11.82）最小的 l 值决定了我们需要让多少个割集（前 l 个）与先前的割集不相交，以达到所需要的精度。

表 11.4 列出了通过使用截断不相交积之和算法得到的 Cm^* 产品的解。为了生成这些割集，Sun-4 型 CPU 运行了 138s。

表 11.4　Cm^* 产品的故障树的解

截断的不相交积之和

（Cm^* 产品的故障树——总共 5405 个割集）

运行时间/h	5	10	100	1000
割集数量 l	2	2	36	80
不可靠度下边界	4.28e^{-7}	1.71e^{-5}	1.84e^{-4}	2.62e^{-2}
不可靠度上边界	4.31e^{-7}	1.74e^{-5}	1.87e^{-4}	2.68e^{-2}
总运行时间（Sun-4 型 CPU，s）	24	24	33	238

11.6.3 Markov 链的截断

当使用 Markov 模型时，最主要的问题是 Markov 链的状态空间可能会根据产品部件的数量呈指数性增加。但是，一个较大 Markov 模型的相对少数状态通常会对所需的结果产生重要影响。假如我们可以提前确定某些状态，且这些状态出现的概率较低，那么我们就可以不用去管这些状态所产生的割集和它们的解。对于那些设计成在相对较短运行时间内提供高可靠度的产品，它们不太可能在较短的运行时间内出现大量的部件失效。因此，多个部件失效的状态出现的概率就会很低。基于这种观测方式的模型简称为状态截断（State Truncation）。

假设在构建 Markov 链的过程中，我们只创建了 k 个部件失效的状态，而把多于 k 个部件失效的状态失效聚合到一个特殊的虚拟状态，我们称之为截断聚合（Truncation Aggregation，TA）状态。一般来说，聚合到 TA 状态的状态包括运行状态和失效状态。如果把 TA 状态看作运行状态，我们就会得到产品可靠度的一个乐观估计值或上边界；把 TA 状态看作失效状态，我们就可以得到产品可靠度下边界的一个保守估计值。

用 Pr（TA）表示 TA 状态的概率，用 Pr（$Down\ States$）表示模型的非聚合部分中（Markov 链截断线之上的部分）失效状态的概率之和。产品不可靠度的边界可用式（11.83）计算：

$$Pr(Down\ States) \leqslant Unreliability \leqslant Pr(Down\ States) + Pr(TA) \tag{11.83}$$

对于 3P2M 产品（案例 11.5、图 11.35），在第一个失效级别之后截断的 Markov 链如图 11.40 所示。除了使用相同的覆盖模型之外，使用和案例 11.5 中相同的参数，此 Markov 链就是产品运行 100h 的解。失效状态的概率之和是 4.52×10^{-4}，它是产品不可靠度的下边界。TA 状态的概率是 6.05×10^{-5}，将它加到下边界后，可以得到不可靠度的上边界 5.12×10^{-4}。

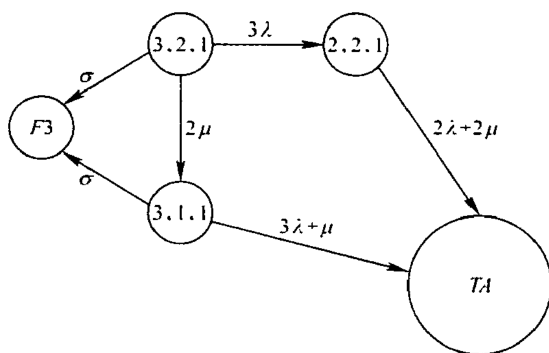


图 11.40　三处理器、二内存系统的截断 Markov 模型

因此，产品可靠度的合理估计可以从整个状态空间中相对较小的子集获得。但是，如果模型在级别 k 上截断所得到的边界太松弛，那么就要对产品重新进行建模，并对其求解，以提高截断所在的失效级别。

11.7 高级主题

这一节将介绍分析容错产品时所要考虑的其他问题，并为读者提供一些可参考文献。

11.7.1 性能与可靠性的结合

在计算机产品中使用多个处理器，可以增强产品性能并提高可靠性。备用处理器能够提高吞吐量，缩减任务的回转时间（Turnaround Time），还能够为产品的关键功能提

供冗余。假设有 n 个处理器的产品，当一个处理器能胜任指定工作，另外 $n-1$ 个处理器作为备用时，产品就可以达到最高的可靠性级别。但是，这种情形也拥有最差的平均利用率，因为许多处理器（备用的）将会在任何给定时间内处于空置状态。提供最佳平均利用率的产品配置应该是在允许非冗余和无最差可靠性的前提下，使用所有 n 个处理器进行计算。在这两种极端情况之间的某处，产品可以同时满足性能和可靠性两个方面的要求。当冗余水平很高的时候，一个逐渐衰退（Gracefully Degradable）的产品仍然可以提供较高的性能表现；只要产品的某些最小配置可用，那么性能就将随部件的失效而降低（但是产品仍可以继续运行）。

分析逐渐衰退产品时，必须同时考虑性能和可靠性 [Meyer, 1992]。利用 Markov 奖励模型（Markov Reward Model）可以轻松完成这项工作。Markov 奖励模型是一个 Markov 链，此模型拥有一个与每个状态相关的额外量度——奖励。这种奖励可能是吞吐量、响应时间、运行开销、计算能力等。在一些简单情况下，奖励量度表现为值为 r_i 的矢量，其中，r_i 是与状态 i 相关的奖励。在 t 时刻的期望奖励 $\{E[R(t)]\}$ 为

$$E[R(t)] = \sum_{i=1}^{n} r_i P_i(t) \tag{11.84}$$

其中，$P_i(t)$ 是产品在 t 时刻处于状态 i 的概率。

令 $r_i=1$ 表示运行状态，$r_i=0$ 表示失效状态，公式（11.84）就可以用来估计产品的可靠度。对于 3P2M 产品（案例 11.5、图 11.35），假设一个处理器、一个内存和总线就可以提供基本性能（$r_{1,1,1}=1$），那么增加第二个内存单元，可以提升产品的性能，性能提升因子为 1.25；增加第二个处理器，也能提升产品的性能，性能提升因子为 1.8；使用 3 个处理器，可以提升基本性能 2.5 倍。表 11.5 列出了运行状态的奖励矢量。3P2M 产品的期望性能是 3.1，略低于与初始状态的峰值性能级别。

表 11.5　3P2M 产品的奖励和状态概率

状　态	奖　励	100h 内的状态概率
3、2、1	3.125	0.98613
2、2、1	2.25	1.33×10^{-2}
3、1、1	2.5	3.18×10^{-5}
1、2、1	1.25	5.99×10^{-5}
1、2、1	1.8	4.30×10^{-7}
1、1、1	1	1.93×10^{-9}

11.7.2 阶段性运行

容错产品经常要在若干阶段内运行。在这些阶段中，产品结构、失效过程或成功运行标准会因阶段的不同而有所变化。例如飞行器控制产品必须在起飞、巡航和着陆阶段内控制飞机。

此类产品的可靠性分析会因为多个运行阶段而受到干扰，因为我们必须为每个阶段构建模型。但问题是每个独立阶段的模型又必须相互连接，所以每个阶段最后的解就变

成了第二个阶段的初始状态。

我们用一个最简单的案例来描述解决阶段性应用问题的一些不同的方法。图 11.41 所示为案例产品的可靠性框图。此产品包含三个部件，它们在三个连续阶段内运行。部件的失效率在每个阶段内恒定但又不同。假如产品在任何运行阶段内失效，产品就会失效。图 11.42 显示了为所有三个阶段构建的 Markov 链模型。Markov 链中的状态的标签是在此状态下仍然运行的部件的名称。贴上 F 标签的状态是失效状态。

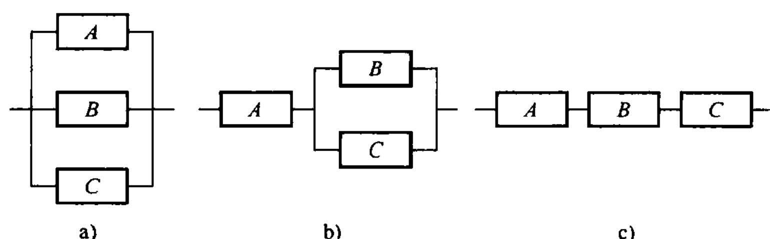


图 11.41　案例系统的三个任务阶段

a）阶段一　b）阶段二　c）阶段三

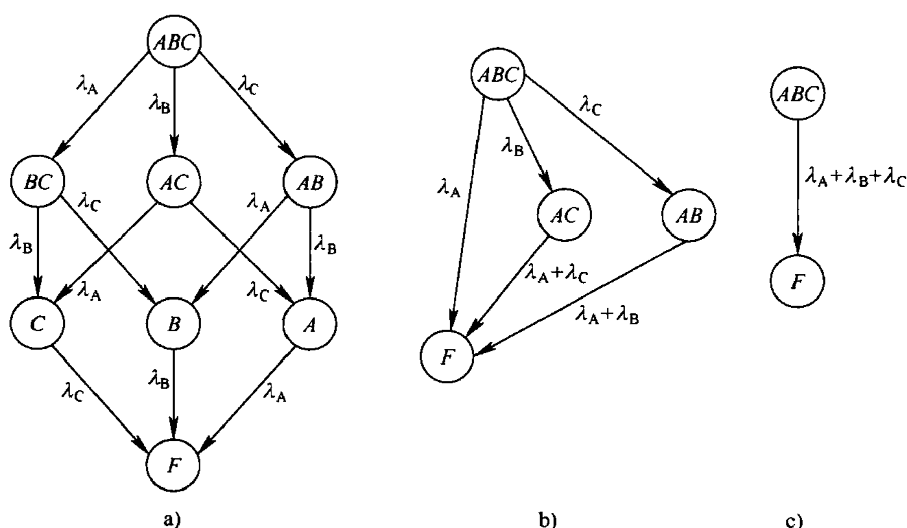


图 11.42　案例系统三阶段的 Markov 模型

a）阶段 1　b）阶段 2　c）阶段 3

阶段性运行分析最简单的方法是把每个阶段的可靠性框图串联起来，然后在最后一个运行阶段计算产品的可靠度。图 11.43 所示为案例产品的可靠性框图。因为在最后阶段的产品运行需要所有三个阶段内的运行，所以此框图可以简化成串联的 A、B 和 C。假如部件是不可修复的，那么此模型的解就是精确的可靠度，但是如果产品运行时，可以对冗余部件进行修复，此模型将产生产品可靠度的保守估计值。为了得到这样的结果，假设部件 C 在阶段 1 中失效，并在阶段 2 中修复。因为 C 在两个阶段中都是冗余的，所以失效/修复过程不会引起产品失效，但是在保守的可靠性模型中，部件 C 失效时，产品也会失效。

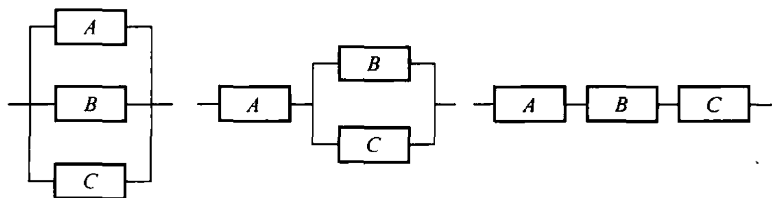


图 11.43 案例的保守可靠性模型

求解阶段运行产品组合解的精确方法会受到用等价产品构建的框架图的影响。在阶段 i 的框图中，部件 C 替换为一系列关联的单个部件 C_1，C_2，…，C_i，如图 11.44 所示。替换后，部件的失效概率是条件概率。部件 C_i 的可靠度是假设 C_i 已经存活到前一阶段，部件 C 在阶段 i 中存活的概率。这种方法可以求出精确解，但要付出一些代价。因为故障树模型的解随着部件数量呈指数性增长，所以替换部件会带来昂贵的计算开销。

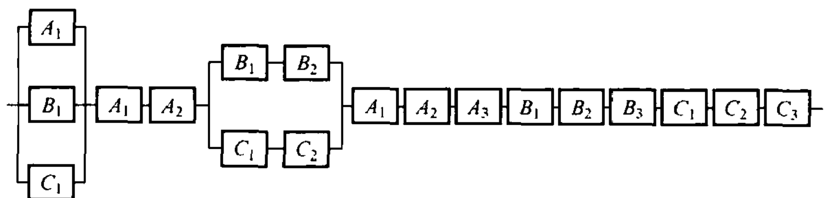


图 11.44 阶段性应用产品的精确可靠性模型

解决阶段运行问题的 Markov 标准方法包括连续求解每个 Markov 模型，把从运行阶段 i 得来的状态概率当作运行阶段 $i+1$ 的初始状态概率。以案例产品为例，假设运行阶段的改变发生在 T_1 和 T_2 时刻，并且最后阶段终止于 T_3 时刻，那么，我们要把从第一阶段模型在 T_1 时刻得来的状态 ABC、AC 和 AB 的概率当作第二阶段的初始状态概率，且使用相同的命名；把其他状态的概率之和当作第二阶段模型中状态 F 的初始状态概率。对于阶段 2 变化到阶段 3，我们需要将状态 ABC 在 T_2 时刻的状态概率用作第三阶段模型中状态 ABC 的初始状概率；剩余状态的概率之和用作失效状态的初始概率之和。状态 ABC 在第三阶段 T_3 时刻的状态概率就是阶段性应用产品的可靠度。

此种方法的一个缺点是除非每个阶段的部件组是相同的，否则将一个阶段模型的状态与另一模型相关状态相匹配是非常困难的。假如部件在第一阶段而不是其他阶段期间失效，或者失效发生在第一阶段却直到后一个阶段才被发现，并在后面的阶段中重新使用，那么这些情况都将产生更多的问题。

文献［Somani and Trivedi，1994］在阶段性应用产品的分析中使用了布尔代数方法，而文献［Smotherman and Zemoudeh，1989］和［Dugan，1991］则使用了 Markov 方法。

11.7.3 高级故障树建模

使用组合模型分析冗余产品时，需要考虑其他一些可用的建模方法。我们为新的方法提供一些文献阅读指导。

① 动态故障树模型。故障树模型的主要限制是它无法捕获序列化的相关性（Sequence Dependency）。当在失效的发生顺序（不是简单的组合）影响产品运行时，将产生序列化

相关结果（Sequence Dependency Result）。除了备用冗余和覆盖建模之外，序列化相关结果包括使用备件共享池（Shared Pools of Spares）和功能关联（Functional Dependencies）。当一些部件由于另一个部件的失效而变得无用或断开连接时，就会出现功能关联。文献[Dugan, Bavuso and Boyd, 1992]介绍了动态故障树，它是使用特殊的逻辑门来处理这种动态行为。文献[Heidtmann, 1992]也介绍了一些分析动态冗余的方法。

② 把覆盖纳入到组合模型中。大多数把容错计算机产品的覆盖纳入到组合模型的技术都要求使用 Markov 模型，但是在近期的一些论文中，介绍了另外一些把覆盖纳入到组合模型的方法[Doyle, Dugan and Patterson-Hine, 1995]。

③ 二进制决策图（Binary Decision Diagrams, BDDs）。故障树模型的解会随部件数量或割集数量呈指数性增长，所以求解过程需要耗费大量时间。大型产品故障树模型具有固有的局限性。近期，有一些用来自电路理论的技术解决故障树模型的方法，即把故障树当作逻辑框图使用。BDD 就是此类方法，它为解决特大故障树的提供了有效途径[Coudert and Madre, 1993]。在许多情况下，使用 BDD 解决故障树的时间与产品割集数量相独立。

④ 软件和硬件的组合模型。一些作者已经考虑在容错计算机产品中对硬件和软件故障进行分析。[Stark, 1987]和[Laprie and Kanoun, 1992]介绍了冗余产品的软件和硬件使用的 Markov 模型。[Dugan and Lyu, 1993, 1994]也曾介绍过使用故障树和 Markov 模型对硬件和软件容错产品进行系统及分析的方法。

11.8　总结

本章介绍了若干种重要的用于分析容错和冗余产品可靠性的技术。我们用诸如可靠性框图和故障树等组合模型分析了静态冗余产品，而 Markov 模型则更适于用来分析动态冗余产品。组合模式具有简明、易懂的优点，但是 Markov 模型更灵活。本章还介绍了时间独立分析以及时间相关度量方法，阐述了不完全覆盖所对预测容错计算机产品的可靠度产生深远影响的可能性，给出了用于分析覆盖的子模型案例以及把覆盖模型纳入到 Markov 产品模型的方法。文献[Dugan and Trivedi, 1989]中有关于把覆盖模型的详细介绍。本章最后介绍了一些高级主题。

参考文献

Coudert, O., and J. C. Madre. 1993. Fault tree analysis: 1020 Prime implicants and beyond.

Proceedings of the Reliability and Maintainability Symposium, January 1993.

Doyle, S. A., and J. B. Dugan. 1995. Combinatorial models and coverage: A bina-ry-decision-diagram (BDD) approach. Proceedings of the Reliability and Maintainability Symposium, January 1995, Atlanta, GA.

Doyle, S. A., J. B. Dugan, and A. Patterson-Hine. 1991. A combinatorial approach to modeling imperfect coverage. IEEE Transactions on Reliability.

Dugan, G. B. 1991. Automated analysis of phased mission reliability. IEEE Transactions on Reliability.

Dugan, G. B., S. Bavuso, and M. Boyd. 1992. Dynamic fault tree models for fault tolerant computer systems. IEEE Transactions on Reliability.

Dugan, J. B., and M. R. Lyu. 1993. System reliability analysis of an N-version programming application. Proceedings of the International Symposium on Software Reliability Engineering, Denver, CO.

. 1994. Dependability modeling for fault tolerant software and systems, ed. M. R. Lyu. New York: John Wiley & Sons.

Dugan, J. B., and K. S. Trivedi. 1989. Coverage modeling for dependability analysis of faulttolerant systems. IEEE Transactions on Computers 38 (6): 775.

Heidtmann, K. D. 1992. Deterministic reliability modeling of dynamic redundancy. IEEE Transactions on Reliability 41 (3), 378—385.

Johnson, A. M., and M. Malek. 1988. Survey of software tools for evaluating reliability, avail-ability, and serviceability. ACM Computing Surveys 20 (4): 227.

Johnson, B. W. 1989. Design and analysis of fault tolerant digital systems. Reading, MA: Addison—Wesley.

Laprie, J. C., and K. Kanoun. 1992. X-ware reliability and availability modeling. IEEE Transactions on Software Engineering 1: 130.

Meyer, J. F. M. 1992. Performability: A retrospective and some pointers to the future. Performance Evaluation 14: 139.

Ng, Y-W., and A. Avizienis. 1976. A model for transient and permanent fault recovery in closed fault tolerant systems. Proceedings IEEE International Symposium on Fault Tolerant Computing, FTCS-6, 182, June 1976.

Rauzy, A. 1993. New algorithm for fault tree analysis. Reliability Engineering and System Safety 40: 203.

Siewiorek, D. P., and R. S. Swarz. 1982. The theory and practice of reliable system design. Bedford, MA: Digital Press.

Smotherman, M. K., and K. Zemoudeh. 1989. A nonhomogeneous Markov model for phased mission reliability analysis. IEEE Transactions on Reliability 38 (5): 585.

Somani, A. K., and K. Trivedi. 1994. Phased-mission system analysis using Boolean algebraic methods. Proceedings of the ACM Sigmetric Conference on Measurement and Computer Systems 98.

Stark, G. E. 1987. Dependability evaluation of integrated hardware/software systems. IEEE Transactions on Reliability 440.

第 12 章　可维修产品的可靠性模型和数据分析

12.1　引言

本章将介绍可维修产品的建模和分析方法。通常，大部分可维修产品，特别是非电子设备都存在磨损现象。大多数可维修产品可通过替换或者修复失效组件的方法恢复到正常状态，而不需要替换或者修复整个单元。修复的质量是核心问题。假如产品在修复之后能恢复到像新产品一样的状况，那么就可以采用普通的修复程序，其失效数据的分析也相对简单。但是如果产品在修复之后无法恢复到像新产品一样的状况，那么失效分析程序以及和分析相关的运行决策都将变得更加复杂。

为了描述可维修产品某些方面的问题，我们举一个例子：假定某汽车已经累计运行30000 英里，但汽车会因为一个小小的轮胎刺孔而发生失效。当轮胎修复后，汽车将重新恢复到可运行状态，但这个小问题的修复对于汽车的寿命产生不了任何重要影响，因此，修复后的产品“像新的一样”，或者是经过更新（Renewal）的设想显然是没有意义的。假如车主决定更换所有四个轮胎，而不是修复刺孔，那么这种部分修复将使产品恢复到新的状况。假如汽车在维修店进行维修时，车主决定大修产品的大部分零部件，那么更新的设想是合理的。

最重要的是，对可维修产品进行分析时，应该考虑修复行为对产品寿命的影响，否则关于运行和维修的政策、备件储备的级别、设计的改进、运输车队的大小以及其他运行和保障要素的决策都将会产生严重错误。第 12.2 节通过引入寿命独立性（Age Independence）和寿命持久性（Age Persistence）的概念为上述问题提供了分析背景。第 12.3 节推荐了一个程序，它用于分析可维修产品的失效数据，并介绍了使用实际数据分析失效的方法。第 12.3.5 节总结了 Weibull 过程数据的分析方法，给出了一个持久寿命过程的特殊案例，它常用于可靠性成长理论（Reliability Growth Theory），此理论对于失效修复过程同样有效。

12.2　分析背景

通过修理能恢复到可运行条件的产品的失效时间呈现为一个随机过程，此过程可表示为$\{X(n);\ n=0, 1, 2, \cdots\}$或简短地表示为$\{X(n)\}$，其中，$X(0)$的值为0。$X(k)$表示以起始时间为参照，第 k 次失效的时间。起始时间表示安装新设备组件或失效产品通过大修恢复到新状态的时间。此过程称为失效-修复（Failure-Repair，F-R）过程。

在开始描述 F-R 过程的类型之前，我们先介绍一些必要的符号：

$F_{X(n)}(x)$——$X(n)$的累积分布函数，即$F_{X(n)}(x)=P[X(n)\leqslant x]$；

$F_{X(n+1)|X(n)=x}(y)$——条件随机变量$X(n+1)|X(n)=x$的累积分布函数，即$F_{X(n+1)|X(n)=x}(y)=P[X(n+1)\leqslant y|X(n)=x]$；

$T(i)$——第i次到达时隔$=X(i)-X(i-1)$，$i=1, 2, \cdots, n$；

$G_n(t)$——T_i的累积分布函数；

$r_x(x)$——由$pdf(x)/cdf(x)$定义的失效函数或失效率函数；

$H(x)$——累积失效率函数；

$N(x,y)$——到达间隔(x,y)内的失效次数；

IFR（DFR）——递增（递减）失效率；

IFRA（DFRA）——递增（递减）失效率的平均值；

IMRL（DMRL）——递增（递减）平均剩余寿命；

NBUE（NWUE）——比使用过产品更好（差）的程度。

在实际中，有两个特征表示维修影响的界限：F-R过程的寿命独立性（Age-Independent）和寿命持久性（Age-Persistent）。

12.2.1 寿命独立F-R过程

F-R过程的第一个特征等同于更新（Renewal）。它表示修复将使设备恢复到新产品一样的状况或恢复到与$T(1)$具有相同失效分布的状况，$T(n)$表示第n次到达时隔（Interarrival Time）。这种维修被称为最大修复，因为它通常涉及所有设备的替换或者大修（Major Overhaul）。术语“寿命独立”或“AI（Age-Independent）”用于描述此类失效修复过程，潜在的失效-影响机制取决于距最近一次维修的时间，而不依赖于以起始状态为参照的设备寿命。从形式上来说，对于$0\leqslant x_n<x_{n+1}<\infty$和所有整数$n$，当且仅当式（12.1）成立时，F-R过程才是AI（寿命独立的）。

$$P[X(n+1)\geqslant x_{n+1}|X(n)=x_n,\cdots]=P[X(1)\geqslant x_{n+1}-x_n] \tag{12.1}$$

因此，假如前一个失效发生在寿命y之后，则在寿命x之后发生第n次失效的概率与新产品在$x-y$小时后发生失效的概率相等。

12.2.2 寿命持久F-R过程

与AI或者最大修复过程相比，最小修复策略假设修复行为使产品刚好恢复到失效之前的可运行状态，所以产品寿命没有变化。这应该是对汽车轮胎刺孔修复的最合理假设。术语“寿命持久（Age Persistence，AP）”用于描述此类过程。对于$0\leqslant x_n<x_{n+1}<\infty$和所有整数$n$，当且仅当式（12.2）成立时，F-R过程才是AP（寿命持久的）

$$P[X(n+1)\geqslant x_{n+1}|X(n)=x_n,\cdots]=P[X(1)\geqslant x_{n+1}|X(1)\geqslant x_n] \tag{12.2}$$

因此，对于一个AP过程，假定产品可以生存到x_n时刻，且第n次失效发生在x_n时刻，那么，第（$n+1$）次失效时间的分布与第一次失效时间的分布相同。

12.2.3 定义AI和AP的特征

根据前面定义的符号，I是正整数集合：

当且仅当$\overline{F}_{X(n+1)|X(n)}(x_{n+1}|x_n)=\overline{F}_{X(1)}(x_{n+1}-x_n)$时，$\{X(n),\ n\in I\}$为AI；

当且仅当$\overline{F}_{X(n+1)|X(n)}(x_{n+1}|x_n)=\dfrac{\overline{F}_{X(1)}(x_{n+1})}{\overline{F}_{X(1)}(x_n)}$时，$\{X(n),n\in I\}$为 AP。

F-R 过程的两种形式以随机变量$X(n+1)|X(n)$的分布为特征，而随机变量则根据第一寿命长度（First Life-Length）的随机变量$X(1)$定义。

如果假设维修不能使产品恢复到比新产品更好的状况，也不能使产品恢复到比失效发生前更糟的状况，那么 AI 和 AP 的特征就是修复质量的边界。对于那些包含组件却易于磨损的产品（例如机械设备），AI 过程是理想的选择，因为这种产品处于性能退化状态，它的寿命将因 AI 修复而被重置为 0。

12.2.4 更新（renewal）过程和 Poisson 过程的失效修复

现在，我们讨论 AI 或 AP 的 F-R 过程与更新（Renewal）和 Poisson 过程之间的关系。

1. 更新过程

假如事件之间的间隔时间拥有相互独立且完全相同的分布（Independently And Identically Distributed，IID），那么生成这一系列事件的过程就称为更新过程。那些失效并用新产品进行替换的同类产品通常就呈现为这种过程。

设F是$X(1)$的潜在分布，同时令$F(n)$为F自身的k重卷积，那么$F(k)$就表示k次 IID 随机变量之和的累积分布函数。$F(k)$是时间在第k次事件上的分布，例如

$$F^{(k)}(x)=P[X(k)\leqslant x] \tag{12.3}$$

又因为

$$P[N(0,x)\geqslant k]=P[X(k)\leqslant x]=F^{(k)}(x) \tag{12.4}$$

所以

$$\begin{aligned}P[N(0,x)=k]&=P[N(0,x)\geqslant k]-P[N(0,x)\geqslant k+1]\\&=F^{(k)}(x)-F^{(k+1)}(x)\end{aligned} \tag{12.5}$$

如果失效分布拥有递增失效率，Poisson 分布提供了$(0,x)$内发生n次或更多失效的概率的上限，即如果F是独立失效修复（Independent Failure Repair，IFR），那么

$$P[N(0,x)\geqslant n]\leqslant\sum_{j=n}^{\infty}\frac{e^{-\frac{n}{\theta_p}}}{j!}\left(\frac{n}{\theta_p}\right)^j \tag{12.6}$$

其中，θ_p是 Poisson 平均到达时间间隔（Poisson Mean Interarrival Time）。

更新函数$M_r(x)$被定义为$(0,x)$内预期的更新（事件）次数，即

$$M_r(x)=E[N(0,x)] \tag{12.7}$$

$M_r(x)$被称为基本更新公式，可表示成一个积分等式，其形式如下：

$$M_r(x)=F(x)+\int_0^x M_r(x-t)\mathrm{d}F(t) \tag{12.8}$$

假如F拥有密度f，那么经微分后，为

$$m_r(x)=f(x)+\int_0^x m_r(x+t)f(t)\mathrm{d}t \tag{12.9}$$

函数$m_r(x)=\mathrm{d}M_r(x)/\mathrm{d}x$被称为更新密度，$m_r(x)\mathrm{d}x$是为间隔$(x,x+\mathrm{d}x)$内更新

（例如 F-R 事件）的无条件概率（Unconditional Probability）或每单元时间的预期更新次数。

基本更新定理表明，每单元时间的预期更新次数接近 $1/\theta_p$，即

$$\lim_{x\to\infty}\left[\frac{M_r(x)}{x}\right]=\frac{1}{\theta_p} \tag{12.10}$$

其中，θ_p 是平均到达时隔。

对于大多数常见分布，在 x 到 $x+h$ 的时间间隔内，预期更新次数近似于 h/θ_p。其中，x 值较大，h 值较小。

根据下面的等式，Laplace 变换可用来求解更新函数或更新密度：

$$M_r^*(s)=\frac{F^*(s)}{1-F^*(s)}$$
$$m_r^*(s)=\frac{F^*(s)}{1-f^*(s)} \tag{12.11}$$

其中，

$$g^*(s)=\int_0^\infty e^{-sx}g(x)\mathrm{d}x \tag{12.12}$$

$g(x)$ 只定义为实线的正部分。

当在间隔（x_1，x_2）内观测到更新过程，则平均到达时隔（即平均失效间隔时间）由式（12.13）计算：

$$E[T(x_1,x_2)]=[M(x_2)-M(x_1)]/(x_2-x_1) \tag{12.13}$$

当及时地在任意点观测时，结果较为准确，假如观察开始于一个更新过程，那么结果接近于真实值。

现在讨论在时刻 x 处开始运行，产品单元剩余寿命的分布情况。假如用 $\tau(x)$ 表示此随机变量，那么

$$P[\tau(x)>y]=\overline{F}(x+y)+\int_0^x\overline{F}(x+y-z)\mathrm{d}M(z) \tag{12.14}$$

注意：在之前的等式中，我们已知从原点 x 开始的时间，但不知道工作中单元的寿命。假如 x 也未知，那么下一个失效的等待时间为

$$W(t)=\frac{1}{\theta}\int_0^t(y)\mathrm{d}y \tag{12.15}$$

更新过程产生的重要结果经常称为 Drenick 定理，它用于处理拥有 n 个独立部件的一系列产品。假如每一个部件的 F-R 过程都是更新过程，那么在相当差的条件下，当部件数量变大时，产品失效间隔时间的极限分布为指数分布；即使部件出现失效的次数不服从指数分布，也会出现这种现象。在可靠性理论中，这个结果相似于中心极限定理，它解释了为什么指数分布能应用于拥有磨损部件的产品。Blumenthal、Greenwood 和 Herbach 在 1973 年的研究也表明：运行周期长度比部件数量更为重要。

更新过程的另一个结果是著名的稳态可用性方程（Steady-State Availability Equa-

tion)。设 T_i 是第 i 次失效出现之前的运行时间，D_i 是第 i 次修复的修复时间。假如 T_i 和 D_i 都表示更新过程，那么，产品在寿命 x 处运行的概率 $A(x)$ 的极限是

$$A=\lim_{x\to\infty}A(x)=\frac{\theta_{ft}}{\theta_{ft}+M_{rt}} \tag{12.16}$$

其中，θ_{ft} 和 M_{rt} 分别表示失效时间和修复时间分布的均值。

2. 齐次 Poisson 过程

齐次 Poisson 过程（Homogeneous Poisson Process，HPP）是一种更新过程，它在任何间隔内的事件数量分布由式（12.17）计算：

$$P[N(x_1,x_2)=m]=\frac{[\lambda(x_2-x_1)]^m}{m!}\mathrm{e}^{-\lambda(x_2-x_1)} \tag{12.17}$$

参数 λ 是时间的倒数，为常数。在较长时间内，它用于测量事件发生率的平均值。因此，我们可以认为 $\lambda(x_1-x_2)$ 是在 (x_1,x_2) 内的平均事件数量。齐次 Poisson 过程的其他特性包括：

① 当且仅当 F-R 过程是 HPP 时，失效间隔时间为指数分布。

② 任意间隔内的事件数量独立于其他非重复间隔中的事件数量；我们可以任意定义起点，并以此为参照来测量失效出现时间。

③ 第 n 次事件的出现时间 $X(n)$ 服从 gamma 分布

$$f_{X(n)}(x)=\frac{\lambda(\lambda x)^{-1}\mathrm{e}^{-\lambda t}}{(n-1)!}(x\geqslant 0) \tag{12.18}$$

同时，$2\lambda X(n)$ 是拥有 $2n$ 自由度的卡方分布。$X(n)$ 的均值和方差分别是

$$E[X(n)]=n/\lambda,\quad Var[X(x)]=n/\lambda^2 \tag{12.19}$$

对于拥有较大值的 n，$X(n)$ 接近于前面给出的均值和方差。

假如有 p 个 HPP 运行（例如设备的 p 个部分，它们具有相同的失效率 λ），同时假定失效与运行过程无关，那么整个过程就是拥有速率参数 $p\lambda$ 的 HPP。假如 p 个设备单元和失效都没有被替换，那么第 r 次失效的分布 $X(r)$ 由式（12.20）计算：

$$X(r)=V_1/p+V_2/(p-1)+\cdots+V_r/(p-r+1),\quad r=1,2,\cdots,p \tag{12.20}$$

其中，V_r 表示拥有参数 λ 的独立指数分布随机变量。因此，

$$E[X(r)]=(1/\lambda)[(1/p+1/(p-1)+\cdots+1/(p-r+1)] \tag{12.21}$$

在间隔 $(0,x)$ 中，满足 $x_1\leqslant x_2\leqslant\cdots\leqslant x_n\leqslant x$ 的 n 次失效的联合分布是

$$f_{x_{(n)}}(x_1,x_2,\cdots,x_n,x)=\lambda\mathrm{e}^{-\lambda x_1}\lambda\mathrm{e}^{-\lambda(x_2-x_1)}\cdots\lambda\mathrm{e}^{-\lambda(x_n-x_{n-1})}\mathrm{e}^{-\lambda(x-x_n)}=\lambda^n\mathrm{e}^{-\lambda x} \tag{12.22}$$

假如在 $(0,x)$ 内已发生 n 次事件，那么 x_1，x_2，…，x_n 的条件概率密度函数（Probability Density Function，PDF）是

$$f_{x_{(n)}}(x_1,x_2,\cdots,x_n|x_n\leqslant x)=\frac{n!}{x^n},\quad 0\leqslant x_1\leqslant x_2\leqslant\cdots\leqslant x_n\leqslant x \tag{12.23}$$

这与在 $(0,x_n)$ 上均匀分布的 n 个随机变量对应的 n 阶统计的分布一样。

3. 非齐次 Poisson 过程

假如 Poisson 过程的发生率是一个时间关联函数（Time-Dependent Function），那么就称它为非齐次 Poisson 过程（Nonhomogeneous Poisson Process，NHPP）。在所有间隔内

发生事件数的概率分布是

$$P[X(x_1,x_2)=m]=\left[\int_{x_1}^{x_2}\nu(x)\mathrm{d}x\right]^m\frac{\mathrm{e}^{-\int_{x_1}^{x_2}\nu(x)\mathrm{d}x}}{m!}\tag{12.24}$$

注意：此概率具有 Poisson 形式 $[\gamma^m!]\ \mathrm{e}^{-\gamma}$。其中，参数 γ 是整数项，它取决于 x_1 和 x_2。函数 $\nu(x)$ 称作强度函数，表示事件发生的时间关联性。我们称 $\nu(x)\Delta x$ 为近似无条件概率，它是在 $(x,x+\Delta x)$ 内发生事件的概率，Δx 为较小值。

$$m(x)=\int_0^x\nu(t)\mathrm{d}t\tag{12.25}$$

式（12.25）称为均值函数（Mean-Value Function），因为它代表从 0 到 x 的预期失效数。

假如对式（12.25）进行时间尺度变换：

$$\tau=\int_0^x\lambda(t)\mathrm{d}t\tag{12.26}$$

那么失效系列将变为齐次 Poisson 过程。对于一个 NHPP 来说，成功运行事件间的间隔是独立分布的。假如我们在 $(0,x_0)$ 内观测运行过程，且事件发生在 $(x_1,x_2,\cdots,x_n)$ 内，那么，观测失效次数的概率函数是：

$$L(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}\nu(x_i)\mathrm{e}^{-\int_0^{x_0}\nu(x)\mathrm{d}x}\tag{12.27}$$

假设在 $(0,x_0)$ 内发生了 n 次事件，那么 n 个失效时间 $(x_1,x_2,\cdots,x_n)$ 的条件 pdf（概率密度函数）与服从常见分布函数的 n 阶统计的条件 pdf（概率密度函数）是一样的

$$F_{x_{(n)}}(x)=\frac{\nu(x)}{\nu(x_0)},0\leqslant x\leqslant x_0\tag{12.28}$$

以上结果与齐次 Poisson 过程的结果相对应。

文献［Balaban and Singpurwalla，1984］得出了大量关于随机变量 $X(n+1)\mid X(n)$ 特性的结果——即已知 NHPP 的第 n 次失效时间，第 $n+1$ 次失效时间。我们将其中一些结果总结如下：

① 当且仅当 $F_{X(n+1)\mid X(n)}$ 为 *IFRA*（*DFRA*）时，$F_{X(1)}$ 为 *IFR*（*DFR*）。上述结果将会导致以下一连串指示关系：

$$\begin{array}{ccc} F_{X(1)}IFR(DFR) & \Leftrightarrow & F_{X(x+1)\mid X(n)}IFRA(DFRA)\\ \updownarrow & & \uparrow\\ F_{X(x+1)\mid X(n)}IFR(DFR) & \leftarrow & \text{- - - - - - - - - -} \end{array}$$

② 当且仅当 $F_{X(n+1)\mid X(n)}$ 为 *NBUE*（*NWUE*）时，$F_{X(1)}$ 为 *DMRL*（*IMRL*）。此结果产生的指示关系链如下：

$$\begin{array}{ccc} F_{X(1)}DMRL(IMRL) & \Leftrightarrow & F_{X(x+1)\mid X(n)}NBUR(NWUE)\\ \updownarrow & & \uparrow\\ F_{X(x+1)\mid X(n)}DMRL(IMRL) & \leftarrow & \text{- - - - - - - - - -} \end{array}$$

③ 对于其他所有失效分布特性，不需要把 $X(1)$ 的值传递给 $X(n+1)\mid X(n)$。

④ 假如 $F_{X(x+1)}$ 为 *IRFA*（*DFRA*），那么，

$$\overline{F}_{X(n+1)|X(n)}(y|x) \leqslant (\geqslant) [\overline{F}_{X(1)}]^{\frac{1}{x+y}} \tag{12.29}$$

⑤ $N_j = X(n)$ 的无条件密度是

$$h_{X(n)}(x) = [H_{X(1)}(x)]^{n-1} \frac{f_{X(1)}(x)}{(n-1)!} \tag{12.30}$$

⑥ 第 n 次失效的时间可以通过最后一次失效时间 x_k 预测，即

$$P[X(n) > x_n \mid X(k) = x_k] = \sum_{m=0}^{n-1-k} e^{-z} \frac{z^m}{m!} \tag{12.31}$$

其中，

$$z = H_{X(1)}(x_n) - H_{X(1)}(x_k) \tag{12.32}$$

4. F-R 过程关系

因为每个失效时间的分布都与新产品的失效时间分布相同，并且所有事件是独立的，所以我们可以将寿命独立过程定义为更新过程。寿命持久失效修复过程的失效时间由 NHPP 决定。文献［Balaban and Singpurwalla, 1984］中的理论已经证明了这个结论。此理论声称：当且仅当 F-R 过程为 NHPP 时，它才是 AP（寿命持久的）。

图 12.1 总结了各种过程之间的关系。这些关系并不能说明只要 $X(1)$ 为指数分布，F-R 进程就是 AP 或 AI，这只是一个必要非充分条件。例如即使伴随修复的失效时间服从指数分布，但只要失效率在两种形式中的变化不一致，那么 F-R 进程就可能既不是 AI 也不是 AP。

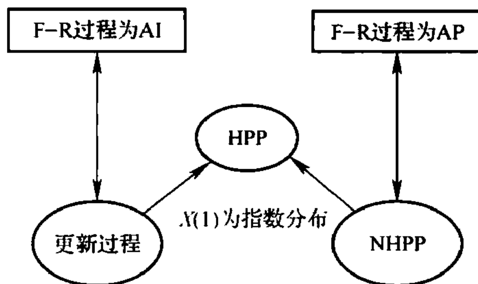


图 12.1　F-R 过程、Poisson 过程和更新过程之间的关系

12.3　数据分析技术

本节主要介绍了用于描述可维修产品可靠性行为的数据分析技术。因此，我们将在有限范围内讨论介绍 F-R 过程的两种类型：AI 和 AP——也就是说，在更新过程和 NHPP 过程为维修效能所限定的界限内进行讨论。

可维修产品失效数据的基本建模和分析策略如图 12.2 所示。首先，必须检验更新过程假设相对于备选假设到达时隔的单调性趋势或检验失效的发生率是单调递增还是递减，这样就可以确定失效间隔时间（到达时隔）是否可以建模为更新过程。假如趋势不明显，那么 HPP 将会是一个适当的模型；假如趋势明显，就要选择 NHPP；假如更新模型和 NHPP 都不合适，那么就必须考虑更复杂的形式或使用非参数方法来建模。

此处仅讨论用于测试趋势的图形化程序、更新过程测试、HPP 测试以及对 Weibull NHPP 的拟合等。本章后面的参考书目介绍了其他建模形式和数据分析方法。

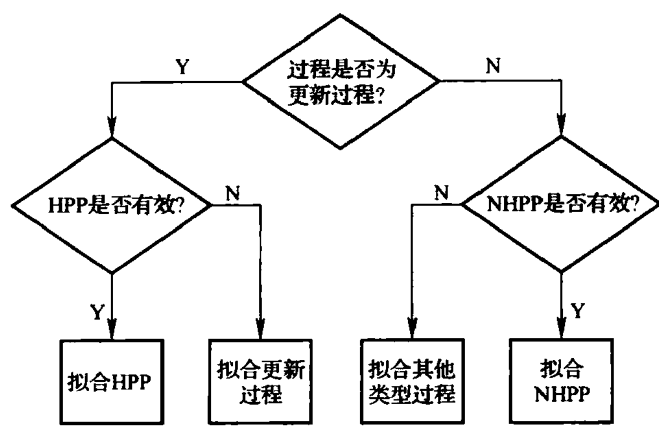


图 12.2　可维修设备的建模策略

12.3.1　图形化趋势测试

图形化程序能为判断失效发生率是否存在趋势提供最初的线索。最简单的图形分析形式是绘制累计失效次数与累计运行时间的关系图。假如图形不存在某种趋势，那么数据拟合线为直线。向上的曲线表示递增的失效趋势。例如随着时间的变化，磨损会频繁导致失效发生。

向下弯曲的曲线表示寿命的改进。图 12.3 显示了所有可能发生的情况。

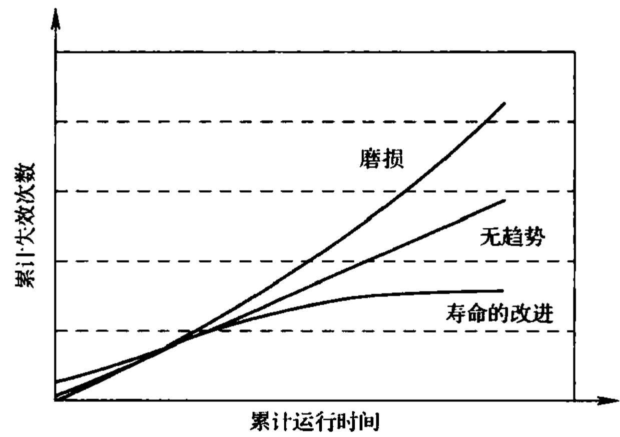


图 12.3　图形化趋势分析

假定某设备单元有以下失效时间：80h、125h、191h、242h、292h、328h、410h、436h、480h、512h、540h、577h、601h、619h、640h、658h、678h、705h、720h 和 741h。累计失效次数与累计运行时间的关系曲线如图 12.4 所示。因为此曲线有一个明确的向上弯曲趋势，所以趋势为失效频率递增。其事实基础是前 10 个到达时隔的平均值是 52.2，最后 10 个到达时隔的平均值是 21.9。

对于单个产品，我们可将数据进行分组并绘制出分组后失效频率曲线。例如以上数据可以分解成 5 个 150h 的间隔，如表 12.1 所示。图 12.5 绘出了分组数据的失效频率，这些数据同样证实了递增趋势这一结论。

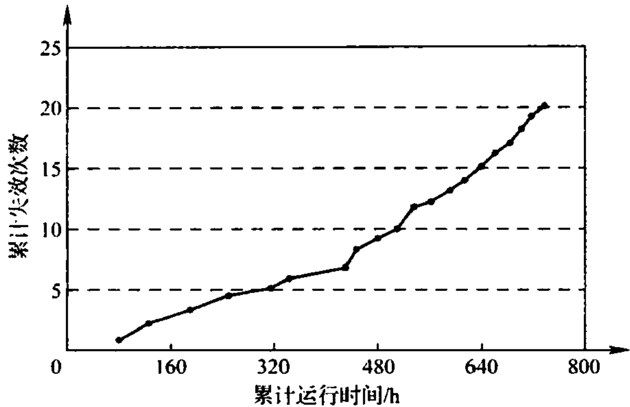


图 12.4 累计失效次数与累计运行时间关系图

表 12.1 失效频率表

间　隔	时间/h	频　率
1	0～150	2
2	150～300	3
3	300～450	3
4	450～600	4

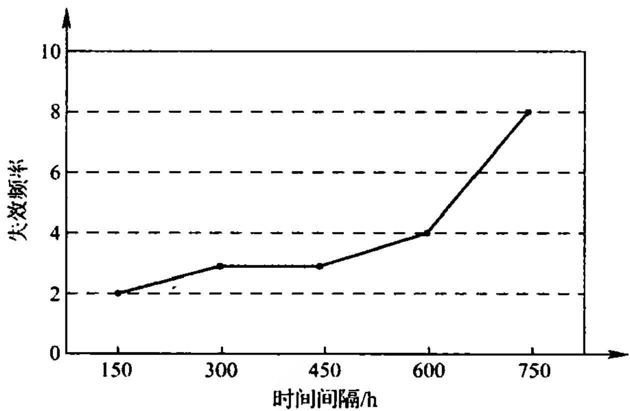


图 12.5 失效频率与时间间隔关系图

此类失效数据可用于大量类似产品。因为观测周期的多样化，我们需要在应用图形化程序之前对数据进行分析。此程序化方法的步骤如下：

① 在观测周期内定义相等时间间隔：$I_1=(0,t)$，$I_2=(t,2t)$，…。

② 定义在每个间隔内所观测产品数量。当观测在一个间隔（审查观测）结束之前停止时，可以使用分数形式表示观测进度。令 n_i 为第 i 个间隔中所观测产品数量。

③ 用等式 $r_t=f_i/r_i$ 表示每一个间隔计算条件失效率，其中，f_i 是在第 i 个时间间隔

内的失效次数。注意：r_i 可大于1.0。

④ r_i 与时间的关系图表示失效发生率与时间之间的一般关系。

如表12.2所示，假定有5个类似单元的观测时间线，其中，每个*表示一个失效发生，T_i 表示第 i 个单元的观测终止时间。通过对这些数据进行分析整理，结果如表12.3所示。注意：样本数量只表示部分在间隔内运行的被检查产品，失效率曲线相对水平，它表示没有明显的失效趋势。

表12.2　失效时间和观测终止时间

产　　品	观测时间/h					
	0~100	100~200	200~300	300~400	400~500	T_i
1	-* *-	-*--	-*--	-* *-	-*--	$T_1=500$
2	-*--	-* *-	-*--	-*--	-*--	$T_2=500$
3	-*--	-*--	-*--	--		$T_3=350$
4	----	-*--	-			$T_4=220$
5	-*--					$T_5=99$

表12.3　失效率估计

时间间隔/h	样本数量	失效次数	失　效　率
0~99	5	5	1.00
100~199	4	5	1.20
200~299	3.2	3	0.94
300~399	2.5	3	1.20
400~499	2	2	1.00

使用此技术处理不同产品的数据时必须谨慎。如果没有明显的证据可以证明产品属于相同类型，那么使用此技术是不合适的。例如有两个产品，其中一个发生失效的5个时间间隔分别是5、4、3、2、1，另一个发生失效的时间间隔是1、2、3、4、5。每组间隔的失效总数都是6，它表示失效发生率是常数。事实上，第一个产品具有较大失效递减率，第二个产品具有较大失效递增率。将二者结合将产生抵消效果，这样可能会给出误导结果。当遇到这类数据时，分析师必须首先确定硬件、运行、环境和数据收集程序的一致性。在下一节中介绍的分析程序将对失效趋势提供更完整的测试。

12.3.2　更新过程测试

更新过程的Mann（1945）测试程序与单调性之间的关系如下：

① 获得时间顺序的到达时隔：

$$
\begin{aligned}
T(1) &= X(1) \\
T(2) &= X(2) - X(1) \\
&\vdots \\
T(n) &= X(n) - X(n-1)
\end{aligned}
\tag{12.33}
$$

② 统计反转 I_n 的数量，即对于每个到达时隔，对它后面的所有到达时隔进行比较，并统计后续到达时隔变大的数量。从数学方面来讲，假如当 $i<j$ 时，$T(i)<T(j)$，则出现反转。

③ 假如 n 小于 10，使用表 12.5 中的值判定更新过程的虚假设是否仍然有效。假如 I_n 的概率级别与所选重要性级别一致，那么就可以把更新过程合理描述为 F-R 过程：对于 $n \geqslant 10$，计算其正态偏差，并与标准正态偏离进行比较：

$$z=\frac{\dfrac{n(n-1)}{4}-I_n}{\left(\dfrac{2n^3-3n^2-5n}{72}\right)^{1/2}} \tag{12.34}$$

表 12.4 中的数据表示某船上水泵单元的失效时间。

$$Z=\frac{10\times\dfrac{9}{4}-10}{\left(\dfrac{2\times10^3-3\times10^2-5\times10}{72}\right)^{0.5}}=2.61 \tag{12.35}$$

表 12.4　泵的失效时间数据

失　　效	失效时间/h	间隔时间/h	反 转 次 数
1	327	327	3
2	1380	1053	0
3	2289	909	0
4	3080	791	0
5	3197	117	2
6	3422	225	1
7	3498	76	2
8	3520	22	2
9	3755	235	0
10	3851	96	—
总计			10

因为 $Z>Z_{0.05}=1.96$，所以与单调趋势相比，更新过程是不适用的。此时的到达时隔以定长递减，并将产生失效发生率递增趋势。

注意：表 12.5 中的最大反转数是 $n(n-1)/2$。对于没有给出的值，使用对称性描述如下：

假如观测反转值 I 的值大于表格中的值，那么

$$\begin{aligned}&\text{如果 } n=8, P(I)=1-P(29-I);\\&\text{如果 } n=9, P(I)=1-P(37-I)。\end{aligned} \tag{12.36}$$

我们使用以下两个例子来阐述对称性：

如果 $n=8$，且 $I=18$，那么 $P(18)=1-P(29-18)=1-P(11)$；
如果 $n=9$，且 $I=20$，那么 $P(20)=1-P(37-20)=1-P(17)$。　(12.37)

假如重要性级别被确定为 P，那么当概率小于等于 $P/2$ 或者大于等于 $1-P/2$ 时，趋势存在，例如10%重要性级别的概率是0.05和0.95。根据所关注的是递增趋势还是递减趋势，我们可以选择使用单尾测试（One-Tail Test）。大量的反转意味着要么失效间隔时间随观测时间递增，要么失效率正在递减。

表12.5　样本量为 n^a 时获取 T 或更小反转的概率

n	T																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	167	500	833	1000																	
4	42	167	375	625	833	958	1000														
5	8	42	117	242	408	592	758	883	958	992	1000										
6	1	8	28	68	136	235	360	500	500	640	765	864	932	962	992	1000					
7	0	1	5	15	35	68	119	191	281	386	500	500	614	719	809	881	932	965	985	995	999
8	0	0	1	3	7	16	31	54	89	138	199	274	360	452	500	500	548	640	726	801	862
9	0	0	0	0	1	3	6	12	22	38	60	90	130	179	238	306	381	460	500	500	540

12.3.3　齐次Poisson过程测试

一个相对简单的测试可用来判定F-R过程是否是HPP（Homogeneous Poisson Process，齐次Poisson过程），它与判断单调趋势所用的方法刚好相反。此测试称为中心极限定理测试（Central Limit Theorem Test）或者Laplace测试，文献［Cox and Lewis，1966］中对此有所介绍。它包括以下两种情况：

情况1：在 X' 时刻停止观测。假设在间隔 $(0,X')$ 中有 n 次失效，分别发生在 X_1，X_2，…，X_n 时刻。计算以下统计量：

$$U=\frac{\sum_{i=1}^{n}X_i-\frac{nX'}{2}}{X'\left(\frac{n}{12}\right)^{1/2}} \tag{12.38}$$

情况2：在第 n 次失效时停止观测。假设 n 次失效分别发生在时刻 X_1，X_2，…，X_n。计算下式：

$$U=\frac{\sum_{i=1}^{n-1}X_i-\frac{(n-1)X_n}{2}}{X_n\left[\frac{(n-1)}{12}\right]^{1/2}} \tag{12.39}$$

为了验证数据不存在某种趋势，且HPP有效是虚假设（Null Hypothesis），在选定的重要性级别 Z_α 上比较统计量 U 和标准正态偏差。假如 U 的绝对值小于 Z_α，我们则不能否决虚假设；假如 $U\geqslant Z_\alpha$，那么失效呈递增趋势；假如 $U\leqslant -Z_\alpha$，则失效呈递减趋势。

将上述等式应用到某单一产品中。假如拥有 m 个近似产品的数据，那么就对收集到的数据进行测试，然后为所有 m 个拥有 HPP 但有不同发生率的产品执行以下测试。

令

$$N_j = \begin{cases} \text{第 } j \text{ 项的失效数(情况 1)} \\ \text{第 } j \text{ 项的失效数} - 1\text{(情况 2)} \end{cases} \tag{12.40}$$

$$X_j = \begin{cases} \text{第 } j \text{ 项的观测周期(情况 1)} \\ \text{第 } j \text{ 项最后一次失效的时间(情况 2)} \end{cases} \tag{12.41}$$

那么，

$$U = \frac{s_1 + s_2 + \cdots + s_m - \frac{1}{2}\sum_{i=1}^{m} N_i X_i}{\left[\frac{1}{12}\sum_{i=1}^{m} N_i X_i^2\right]^{1/2}} \tag{12.42}$$

其中，

$$s_j = \sum_{i=1}^{N_i} X_{ij} \tag{12.43}$$

s_j 表示第 j 个产品的失效时间之和。

为了证明计算的合理性，假设产品的观测失效时间分别是第 70h、122h、1525h 和 170h。观测在最后一次失效时停止，此产品属于单一产品，此属于情况 2。因此，

$$U = \frac{\sum_{i=1}^{n-1} X_i - \frac{(n-1)X_n}{2}}{X_n\left[\frac{(n-1)}{12}\right]^{1/2}} = \frac{70 + 122 + 152 + 165 - 4(170/2)}{170(4/12)^{1/2}} = 1.72 \tag{12.44}$$

U 值为 1.72，表示其重要性级别在 10% 上，存在一个良好的失效趋势。

12.3.4 两样本的比较

在很多情况下，所建立的两个样本是否来自于同一总体非常重要，因为此类测试有可能要用作判定是否发生了设计更改、大修是否具有经济效益、运行环境是否已经改变等。

Mann-Whitney 测试程序如下：

① 设 $T_1(1)$，$T_1(2)$，…，$T_1(k)$ 为样本 1 的到达时隔，$T_2(1)$，$T_2(2)$，…，$T_2(m)$ 为样本 2 的到达时隔。

② 将 $k+m$ 个时隔按照从小到大的顺序排列，并设 1 为最小级别，2 次之，依次类推。最大级别的时隔为 $k+m$。假如两个或更多时隔的级别相同，那么使用并列排序。

③ 设 S 是样本 1 的级别之和。也就是说，假如 $R(T_1(i))$ 是第一个产品第 i 次失效时间的级别，那么

$$S = \sum_{i=1}^{k} R[T_1(i)] \tag{12.45}$$

④ 计算测试统计量 U：

$$U = S - k(k+1)/2 \tag{12.46}$$

⑤ 文献［Conover，1983］曾描述到：k 或者 m 小于 8 时，将 U 的临界值与测试统计量 U 进行比较，可以判定两样本来自同一总体的虚假设是否成立。假如 U 大于两个给定临界值中的较大值，那么结论是总体 1 比总体 2 更具可靠性；当 U 小于最小临界值时，结论是总体 1 的可靠性较低。

⑥ 假如 k 和 m 都等于或大于 8，那么可以使用正态逼近法，计算如下：

$$Z=\frac{U-\frac{1}{2}km}{[km(k+m+1)/12]^{1/2}} \tag{12.47}$$

同时，将其与选定的重要性级别的标准正常值比较。

为了便于描述，假定某运行船舶设备在大修前后的数据如表 12.6 所示，表中的到达时隔数据与表 12.7 中所列相似，所有时隔排序后的结果如表 12.8 所示。

将“大修前”的级别相加，可以得出：

$$S=1+3+6+8+9+10=37 \tag{12.48}$$

表 12.6　到达时隔数据

大修前/h	大修后/h
327	161
1380	249
2289	590
3080	901
3197	
3261	

表 12.7　到达时隔数据（排序后）

大修前/h	大修后/h
327	161
1053	88
909	341
791	311
117	
64	

表 12.8　到达时隔数据级别

到达时隔/h	级　别	样本来源
64	1	大修前
88	2	大修后
117	3	大修前
161	4	大修后
311	5	大修前

（续）

到达时隔/h	级　别	样本来源
327	6	大修后
341	7	大修前
791	8	大修后
909	9	大修前
1053	10	大修后

$k=6$ 时的测试统计量为

$$U=37-6(7/2)=16 \tag{12.49}$$

在 10% 级别上的临界值是 4 和 26 [Conover，1983]。因为 $U=16$ 在此范围内，所以得出的结论是：产品在大修之后的可靠性没有改变。

12.3.5 Weibull 非齐次 Poisson 过程的拟合

如果可以用第 12.3.3 节中所介绍的趋势测试来确立一个失效趋势，那么下一步就要观察是否可以把数据模拟为非齐次 Poisson 过程，是否可以为过程建立等式，以便估计诸如平均失效间隔时间、平均失效次数和失效概率等相关特征。如果过程是非齐次的，那么失效发生率与时间有关；这些特征也会随时间改变，这与 HPP 不同。如前所述，HPP 的属性定义为寿命从 0 到 x 的失效数分布，即

$$P[(0,x)\text{内的失效数}]=\frac{e^{-R(x)}[R(x)^{n}]}{m!} \tag{12.50}$$

其中，$R(x)=\int_0^t r(t)\mathrm{d}t$；

$r(t)$ 是表示瞬时失效发生率的强度函数。在下述情况中

$$r(t)=\lambda\beta t^{\beta-1} \tag{12.51}$$

NHPP 被称为 Weibull 过程。人们已经为此过程提出了大量理论，它也可以作为 NHPP 的模型使用。

1. Weibull 过程的特征

Weibull 过程的特征包括：

① 累积 MTBF$(0,x)$：

$$\theta(0,x)=\frac{x^{1-\beta}}{\lambda} \tag{12.52}$$

② 在寿命 x 处的瞬时 MTBF：

$$M(x)=\frac{x^{1-\beta}}{\lambda\beta} \tag{12.53}$$

③ 在时刻 t 的失效时间分布，t 来自于 x：

$$F_x(t)=1-e^{[illegible]} \tag{12.54}$$

④ 在 (x_1,x_2) 内的预期失效次数：

$$E[N(x_1,x_2)]=\lambda(x_2^{\beta}-x_1^{\beta}) \tag{12.55}$$

⑤ 在(x_1,x_2)内的失效次数分布：

$$P[N(x_1,x_2)=n]=\frac{1}{n!}[\lambda(x_2^{\beta}-x_1^{\beta})]^n\mathrm{e}^{-\lambda(x_2^{\beta}-x_1^{\beta})} \tag{12.56}$$

为了阐述这些等式的使用过程，设$\lambda=0.1$、$\beta=0.2$，时间以月为单位计算，则

① 0至6个月中的累积MTBF是

$$\theta(0,6)=(6^{1-2}/0.1)\text{月}=1.67\text{月} \tag{12.57}$$

② 12个月中的瞬时MTBF是

$$M(12)=[12^{1-2}/(0.1\times2)]\text{月}=0.42\text{月} \tag{12.58}$$

③ 服务寿命为12个月的产品在下半个月内的失效概率是：

$$F_{12}(0.5)=1-\mathrm{e}^{-0.5/42}=0.304 \tag{12.59}$$

④ 从6个月到12个月的预期失效数是：

$$E[N(6,12)]=0.112^2-6^2\mathrm{e}^{-0.1(12^2-6^2)}=10.8\text{次失效} \tag{12.60}$$

⑤ 在6个月到12个月之间，发生8次失效的概率是：

$$P[N(6,12)=8]=\frac{1}{8!}[0.1(12^2-6^2)]^8\mathrm{e}^{-0.1(12^2-6^2)}=0.094 \tag{12.61}$$

2. λ和β的估算

我们必须考虑收集某些数据的可能性。如果数据未分组（U），那么就可得到每个产品每次失效的时间；如果数据已分组（G），这样就只能得到固定间隔内的失效总数。对于未分组数据，观测可以止于某些给定时刻（T）（时间截尾，Time-Truncated），或者止于具体指定数量失效出现的时刻（N）。这些可能性将导致以下三种情况：

①（U-T）未分组，时间截尾。

②（U-N）未分组，失效截尾。

③（G）已分组。

1）对于情况U-T的估算。为未分组数据进行时间截尾测试——在已知（0，x）上n次失效时间的情况下，为β进行最大似然估算（Maximum Likelihood Estimate，MLE），即

$$\hat{\beta}=\frac{n}{n\ln x'-\sum_{i=1}^{n}\ln x_i} \tag{12.62}$$

对于λ，MLE为

$$\lambda=\frac{n}{(x')^{\hat{\beta}}} \tag{12.63}$$

2）对于情况U-N的估算。为未分组数据基于失效次数n的失效截尾测试，即

$$\hat{\beta}=\frac{n}{(n-1)\ln x_n-\sum_{i=1}^{n-1}\ln x_i}$$

$$\hat{\lambda}=\frac{n}{x_n^{\hat{\beta}}} \tag{12.64}$$

3）对于情况 G 的估算。已分组数据的估算过程有些复杂，因为它没有计算 β 时所使用的封闭式方程。假定有 k 个间隔时间的边界是 $x_0=0$，x_1，x_2，…，x_k。那么，β 可以用式（12.65）的解来估算：

$$\sum_{i=1}^{n} n_i \frac{x_i^{\hat{\beta}}\ln x_i - x_{i-1}{}^{\hat{\beta}}\ln x_{i-1}}{x_i^{\hat{\beta}} - x_{i-1}{}^{\hat{\beta}}} - \ln x_k = 0 \tag{12.65}$$

其中，x_0/nx_0 被定义为 0，必须用数值方法求解此等式来获得 β。对于给定的 β 估计值，λ 可以由式（12.66）估算：

$$\hat{\lambda} = \frac{\sum_{i=1}^{k} n_i}{x_k^{\beta}} \tag{12.66}$$

案例 12.1

使用第 12.3.2 节中水泵的数据。前面的讨论表明，它具有单调递增趋势。假如观测停止于 4162h，数据可根据时间截尾试验进行分组，所以，此情况为 U-T。因此，

$$\hat{\beta}\frac{n}{n\ln x' - \sum_{i=1}^{n}\ln x_{(i)}} = \frac{10}{10\times 8.3338 - 77.8094} = 1.81 \tag{12.67}$$

那么，

$$\hat{\lambda} = \frac{n}{(x')^{\hat{\beta}}} = \frac{10}{4162^{1.81}} = 2.8\times 10^{-6} \tag{12.68}$$

注意：β 的估计值大于 1.0，这与失效发生率的递增趋势一致。

3. 拟合优度测试

拟合优度测试可用于观测失效数据是否与 Weibull 过程具有一致性。一般来说，为了应用拟合优度测试，至少应该观测 20 个失效时间。下面将逐一介绍以上三种情况所用的公式。

1）应用于情况 U-T 的测试。计算：

$$G_{\mathrm{UT}} = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{x_i{}^{\bar{\beta}}}{T} - \frac{2i-1}{2n}\right]^2 \tag{12.69}$$

其中，

$$\bar{\beta} = \frac{(n-1)}{n}\hat{\beta} \tag{12.70}$$

将 G_{UT} 与 Cramer-Von Mises 测试的临界值进行比较。假如 G_{UT} 超过了已选定重要性级别值，那么数据与 Weibull 过程一致的虚假设将被否决。

2）应用于情况 U-N 的测试。计算：

$$G_{\mathrm{UN}} = \frac{1}{12(n-1)} + \sum_{i=1}^{n-1}\left[\left(\frac{x_{(i)}}{T}\right)^{\bar{\beta}} - \frac{2i-1}{2(n-1)}\right]^2 \tag{12.71}$$

其中，

$$\bar{\beta} = \frac{(n-1)}{n}\hat{\beta} \tag{12.72}$$

3）应用于情况 G 的测试。为每个间隔计算预期失效数：

$$e_i=\hat{\lambda}(x_i^{\hat{\beta}}-x_{i-1}{}^{\hat{\beta}}),i=1,2,\cdots,k \tag{12.73}$$

如果需要的话，合并相邻间隔时间，这样就可使预期失效数至少为5。如果这样分组之后，有 k' 个间隔时间。令 n' 为调整后第 i 个间隔时间内的失效次数，e' 为相应的预期失效数，然后计算：

$$C=\sum_{i=1}^{k}\frac{(x_i'-e_i')^2}{e_i'} \tag{12.74}$$

C 的分布与二自由度卡方检验的 k' 分布相似，边界值可以在卡方分布表中找到。

4. 置信区间估计

此小节介绍用于计算未分组数据 Weibull 特征的置信界限。在某些情况下，这些界限是近似值，例如两种截尾方式之间的界限就没有区别。大部分界限具有形式为 $C=A(n-1)/n$ 和 $D=B(n-1)/n$ 的因子，其中，n 是失效次数，A 和 B 取决于置信水平和失效次数 n，C 和 D 在文献［Crow，1975］中以表格形式列出。当 $n>60$ 时，可以用式（12.75）对其进行近似计算：

$$\begin{aligned}C&=\left[1-\left(\frac{2}{n}\right)^{1/2}X_{\alpha/2}\right](n-1)/n\\D&=\left[1-\left(\frac{2}{n}\right)^{1/2}X_{\alpha/2}\right](n-1)/n\end{aligned} \tag{12.75}$$

其中，$Z_{\alpha/2}$ 是标准正态分布 $1-\alpha/2$ 百分位处的值。

我们可以使用下面的置信度公式进行计算：

① 强度函数 $r(x)$：

$$\begin{aligned}&\text{LCL:}\quad \hat{e}_L(x)=C\,\hat{r}(x)\\&\text{UCL:}\quad \hat{e}_U(x)=D\,\hat{r}(x)\end{aligned} \tag{12.76}$$

② 预期失效次数 $N(x_1,x_2)$：

$$\begin{aligned}&\text{LCL:}\quad N_L(x_1,x_2)=C\,\hat{\lambda}(x_2^{\hat{\beta}}-x_1^{\hat{\beta}})\\&\text{UCL:}\quad N_L(x_1,x_2)=D\,\hat{\lambda}(x_2^{\hat{\beta}}-x_1^{\hat{\beta}})\end{aligned} \tag{12.77}$$

③ 累积 MTBF，$\theta(0,x)$：

$$\begin{aligned}&\text{LCL:}\quad \theta_L(0,x)=\frac{x}{N_U(0,x)}\\&\text{UCL:}\quad \theta_U(0,x)=\frac{x}{N_L(0,x)}\end{aligned} \tag{12.78}$$

④ 瞬时 MTBF，$M(x)$：

$$\begin{aligned}&\text{LCL:}\quad M_L(x)=\frac{1}{\hat{r}_U(x)}\\&\text{UCL:}\quad M_L(x)=\frac{1}{\hat{r}_L(x)}\end{aligned} \tag{12.79}$$

其中，LCL 是置信下限，UCL 是置信上限。

目前，还没有已分组数据可用的置信度方程。一个保守的方法是把分组数作为前一等式中的 n 使用，这样得到的置信界限将比实际值宽。

12.4 总结

本章阐述了用于分析可维修产品可靠性的不同方法。首先引入了寿命独立和寿命持久概念，它们用于定义维修行为所影响的边界，同时并将它们与著名的更新过程和 Poisson 过程关联了起来；然后又详细阐述了产品失效数据建模和分析的基本策略；还介绍了用于判断失效发生率趋势的图形化和分析测试方法；最后为 Weibull 非齐次 Poisson 过程提出并阐述了详细的优度拟合和估计程序。

参考文献

Balaban, H., and N. Singpurwalla. 1984. Stochastic properties of a sequence of interfailure times under minimal repair and under revival. In Reliability theory and models, ed. M. Abdel - Hameed, E. Cinlar, and J. Quinn. New York: Academic Press.

Blumenthal, S., J. Greenwood, and L. Herbach. 1973. The transient reliability behavior of series systems on superimposed renewal processes. Technometrics 15: 255.

Conover, W. 1971. Practical nonparametric statistics. New York: John Wiley & Sons.

Cox, D. R., and P. A. W. Lewis. 1966. The statistical analysis of series of events. London: Methuen.

Crow, L. H. 1975. Tracking reliability growth. U. S. Army Materiel Systems Analysis Agency, Aberdeen Proving Grounds, MD.

Mann, H. B. 1945. Nonparametric tests against trend. Econometrika 13: 245.

第 13 章　持续的可靠性改进

13.1　引言

可靠性改进技术（Reliability Improvement Technique）可用于各种产品开发情况，例如它可用于那些主要硬件和（或）软件已经通过设计评审的新产品；可用于厂家希望提高其竞争力的处于开发阶段的产品；也可用于一个不符合顾客可靠性要求的现有产品。按道理来说，最后一种情况是不应该发生的，因为在一个产品完成设计并投入批量生产前，对其可靠性水平的要求就应该已纳入到产品的设计中。可靠性改进流程可以发现改进一个复杂产品可靠性的机会，并为这些改进分配时间。让产品在以某种方式工作或对其进行试验，将会暴露那些由设计、制造和（或）操作导致的产品缺陷，这样才能采取措施以消除这些缺陷，也能重新评估，确保并改进产品设计具有可靠性所采用的方法。相比之下，可靠性鉴定试验是为了验证产品在预期环境下工作的能力，而环境应力筛选试验的目的只是暴露缺陷。这些方法本身并不会改进产品的可靠性。

在整个生命周期内，持续的可靠性改进方案将提高产品所能带来的经济效益。通过减少保修期内的维修、维护次数和备件的数量，可以提升商业产品的经济效益。

本章将讨论可靠性增长、加速试验和持续可靠性改进项目管理的一些原则。

13.2　可靠性的增长过程

当在一个复杂设备的设计中使用了创新技术或先进的生产方法，那么这个设备往往有一些无法预料的设计、生产或运行方面的缺陷，这些缺陷影响着设备的可靠性。可靠性改进计划旨在通过改进产品设计实现可靠性目标。改进方案的目标是为了识别、定位和纠正那些存在于设计、制造和运行中的缺陷和薄弱环节。

可靠性改进往往是通过一个以试验、分析和改进（Test, Analyze, And Fix, TAAF）为理念的计划来实现的。当那些消除产品设计缺陷和薄弱环节的纠正措施得以实施，其实施结果在进一步试验中得到验证后，产品的可靠性就得到了改进。TAAF 流程不仅可以应用到新产品的开发中，也可以应用到那些已经投入现场使用的产品。TAAF 流程适用于实验性的或处于样机阶段的产品设备，应用过程可描述为一个反馈循环，如图 13.1 所示。使用反馈循环是可靠性改进计划成功的基础。

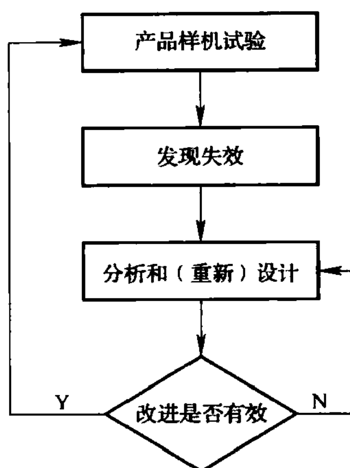


图 13.1 试验、分析和改进（TAAF）流程

13.2.1 可靠性改进计划

可靠性改进计划的试验时间长短取决于以下几个方面：商业产品已知或期望的可靠性水平；军事产品的可靠性要求；产品或样机的成熟度和复杂性；可用于试验的产品数量（试验样品）；失效报告、分析和纠正措施计划的效率以及可用的总试验时间等。由于需要时间进行失效检修、分析已发现失效、研究纠正措施、执行更改，因此，试验时间只是总可用时间的一部分。项目规划和有效的培训可以避免其他一些影响试验时间的因素，这些因素包括操作失误、试验环境故障、试验备件不足和监管不力等。

试验时间除以总可用时间（总可用时间等于规划日历时间乘以试验样品数量）称为试验效率（Test Efficiency）。经验表明，大多数改进计划的试验效率只有 50%，甚至更低。如果增加失效分析的人手，提供充足的试验备件，并进行充分的规划与培训，试验效率也可能高于 50%。然而，接近于 90% ~100% 的试验效率通常表明项目没有改进产品的可靠性。环境应力不够苛刻而不能暴露故障、对可靠性改进计划的管理不够充分、对失效检测和报告系统的不足都有可能导致计划不能改进产品的可靠性。

如果日历时间（Calendar Time）有限，就必须增加试验样品的数量或采用加速寿命试验（见第 13.3 节）。当项目所需的资源和资金充足时，必须通过增加试验样品数量，以尽量减少试验的日历时间。当有多个试验样品时，应为各试验单元或样机合理地分配试验时间。合理分配试验时间，可以防止当“真正的”样品由于失效而停机进行维修时，把时间只花费在那些“镀金”或“手工制作”样品的试验上。仅对一个手工制作的样机进行试验，可能只会达到“通过试验”的效果。为了防止试验的偏差，必须对每个样品进行至少 75% 的平均试验时间的试验。例如对于一个 2000h、两个样品的试验，每个样品都必须至少进行 1000 ×0.75h，即 750h 的试验。

对于那些可靠性不是用单位时间失效数来衡量的产品，也可以对其实施可靠性改进计划。除了以时间为单位外，我们也可以用一些其他的单位用来衡量样品的试验周期，如（汽车）失效前行驶的里程数、（飞机）失效前的飞行次数、（复印机）失效前复印

的张数等。本章讨论以试验时间为单位的可靠性增长和试验持续时间。

在产品进行正式试验前，必须先完成的一项工作是环境应力筛选（Environmental Stress Screening，ESS）。ESS是一种用随机振动和温度循环暴露不良零件和工艺缺陷的制造方法。ESS可以确定产品的早期失效，如制造缺陷（零部件缺失、误用、混用）、工艺缺陷（虚焊、弯脚、连接脆弱）和错误的零部件类型。实施消除这些缺陷的纠正措施，可以降低生产、返工和产品生命周期的成本。ESS可以应用于任何子系统或最终产品，然而只有将ESS应用于最低级别的产品组件时，它才会发挥最大的经济效益。

ESS中的最高和最低温度值都不应超过任何组成产品的零部件或材料的额定值。应当谨慎选择ESS的温度范围，一方面要确保载荷够大而有效激发零部件或材料失效机理的相关物理反应，从而达到有效的筛选；另一方面载荷的范围不应该超过该产品的固有承受能力。考查更高应力水平试验产生的失效或结果，可以持续监测筛选的效力。如果这些试验暴露的失效是因工艺或生产导致的，那么就必须调整筛选的应力。当ESS产生或暴露的失效数据可以接受时，即当绝大多数的失效都是由设计而引发的时候，就可以开展正式的可靠性增长试验了。用于可靠性增长试验的产品必须先接受ESS的考验。

除了ESS之外，还需要完成以下五个方面的工作：

① 确认模拟现场环境试验设备的性能（用来对样机进行试验的温度箱、振动台以及试验测量设备）。

② 完成该产品的热分析。

③ 完成失效模式和影响分析（FMEA）。

④ 建立一个闭环失效报告、分析和纠正措施系统（Failure Reporting，Analysis，And Corrective Action System，FRACAS）。利用FRACAS对在试验中发生的所有失效进行分析，并执行纠正措施，这些失效不仅仅包括发生在正式改进项目中的失效，还包括发生在ESS中的失效。

⑤ 完成可靠性改进计划。

失效分析和纠正措施是可靠性改进计划中至关重要的部分。必须从根源失效模式上解决发生的失效。电气产品的常见失效模式包括焊点裂纹、电路板分层、元器件失效、软件错误、程序错误、电路板布置缺陷和制造工艺问题等。机械产品的常见失效模式包括腐蚀、粘连和破裂（裂纹扩展）等。必须查明具体的失效根源，如过温、电气过应力、杂质沾污、磨损和机械损伤等。一个准确而完整的FMEA分析过程将有助于节省宝贵的时间。零部件失效分析的基本步骤如下：

① 确认使用了正确的零部件。

② 收集零部件历史记录，以确定以往发生了哪些失效及失效的原因。

③ 证实发生的失效。

④ 按照有序的失效分析流程对失效零部件进行分析，如图13.2所示。

⑤ 确定失效模式及其原因。

⑥ 进行电镜扫描和X射线分析。

⑦ 采取推荐的纠正措施，防止相同的失效再次发生。

⑧ 完成一份简洁的报告，总结每一个分析步骤。

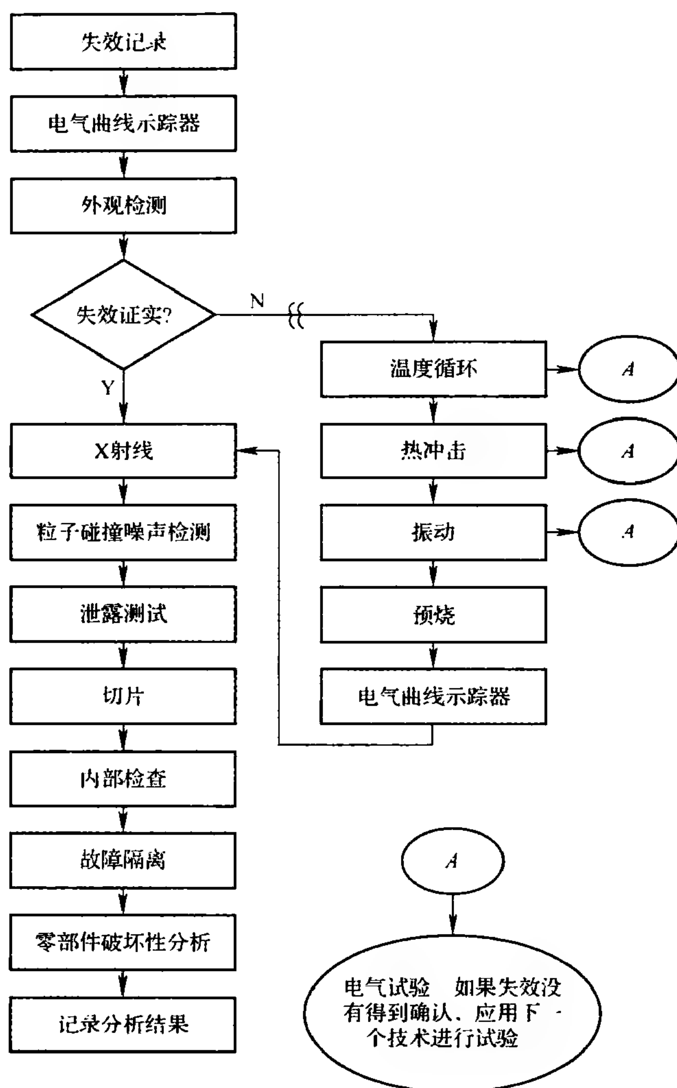


图 13.2　失效分析流程

图 13.2 总结了确定电子零部件失效根源的步骤。一旦确认了零部件，并确定了以往的失效模式和原因后，就应该检查部件外观是否有过应力迹象，然后就可以使用曲线示踪器进行电气测试。如果确认了失效，就可以进一步使用一些无损分析技术，如 X 射线、粒子碰撞噪声检测（Particle Impact Noise Detection，PIND）或泄漏测试（Leak Test）来分析失效的根本原因，接着对部件进行切片或者分解并进行内部检查。采用电子显微镜扫描、激光探测或能量扩散 X 射线分析等技术进一步隔离失效。破坏性键合拉力试验（Destructive Bond Pull Test）可以分析键合失效（Bond Failure）。如果失效没有得到证实，就需要对这些未经证实的失效进行进一步的试验（如热循环、热冲击及振动），直至暴露出失效。即使经历了这些附加试验，有些失效模式可能仍然无法得到

证实。如果发生这种情况，管理层将不得不考虑投入额外资源（如人力资源和成本等）来尝试证实这些问题。

一旦完成了失效分析、确定了纠正措施并记录了分析结果，就需要将这些信息输入到FRACAS中。制造商将利用FRACAS中的这些信息，对产品执行纠正措施。为确保FRACAS的有效性，必须把它纳入到可靠性改进计划和程序中。

可靠性改进计划必须是完整的、经过批准的，只有经过相关责任人的协商和调整才能完成。相关人员包括试验工程师、设计工程师、可靠性经理、生产经理、物流经理和项目经理等。对于军用合同或一些商用合同（如提高某电厂变压器的可靠性），客户应确保计划和程序综合考虑了生产活动和用户代表的意见。计划应至少阐述试验时间安排、所需资源、试验设备、人力、试验环境、试验程序、与试验时间对应的增长计划、产品失效报告以及采取纠正措施的计划等内容。为了确保可靠性改进计划的成功，该计划和程序必须充分描述试验的所有方面，包括一些基本规则。建立试验的基本规则和失效分类准则是可靠性增长试验成功的关键。

13.2.2 失效分类

在可靠性改进计划中，如果没有资源、资金和时间的限制，人们往往会集中精力于设计缺陷的鉴别和排除，而不是失效的分类。在理想的情况下，我们可以纠正那些已经得到确定的失效根源。然而，由于客户预算、进度要求或试验的自身特点［需要通过对失效进行计数和（或）分类，以监控试验的进展］的限制，这种理想的情况往往是不存在的。在一些项目合同中，生产商指定的某些失效类型可能是非关联（Nonrelevant）失效或非责任（Nonchargeable）失效，这可以避免在调查这些失效的根源上花费巨资。然而，客户可能会质疑这些决定，要求查明问题并进行纠正。要尽量避免发生这种对立的状况。

为了确保可靠性改进计划的成功，我们首先应该把所有的失效都认为是关联失效（Relevant Failure），然后分析每个硬件和软件的失效，包括由试验电缆松动（通常是由于间歇性失效或非重现性失效导致的）、试验设备故障或其他试验设备造成的问题等，并为它们制定相应的纠正措施。当所有参与者都注意到调查所有失效带来的益处时，关于失效的分类方法将产生争议。为了尽量减少这种问题的发生，应该在增长试验开始之前建立关于失效分类的基本规则。

图13.3是一个标准的失效分类方法。任何试验设备的异常表现都可以划分或评定为关联失效或非关联失效（有些合同还规定必须确定所有失效的根本原因）。任何一个不会出现在产品实际运行中的异常都被界定为非关联失效。非关联失效往往是由于安装不当、意外损坏、处理不当、试验设备故障或超过试验额定载荷的外部应力而引起的。要判断失效是否为非关联失效，首先必须清楚产品的强度和应力分布（见第13.3节）［Seusy，1987］。理解强度及其分布有助于判断产品是否能够在失效条件下成功地运行。

关联失效（Relevant Failure）包括所有非关联失效（Nonrelevant Failure）之外的失效，不论这些失效是否经过验证。例如间歇性失效，即设备功能的瞬间停止，它就是相关失效。排除故障时出现的间歇性失效也可以归类为关联失效。必须对所有关联失效进

行分析调查，它们可能会导致设计或生产的更改。

任何一个被列为关联失效的异常可进一步分类为责任（Chargeable）失效或非责任（Nonchargeable）失效。非责任失效是由相依失效（Dependent Failure）引起的另一种失效，这类失效是由政府或客户提供的设备导致的或者零部件超过其指定寿命而引起的。责任失效包括间歇性失效，非设备设计性失效，设备及其零部件制造失效，零件设计失效，承包商提供的设备（Contractor-Furnished Equipment，CFE）和承包商的运行、维护或维修程序造成的失效等。具有相同失效原因、失效模式和失效环境条件的所有失效仅算作一次责任失效。责任失效可作为跟踪可靠性增长的基础。

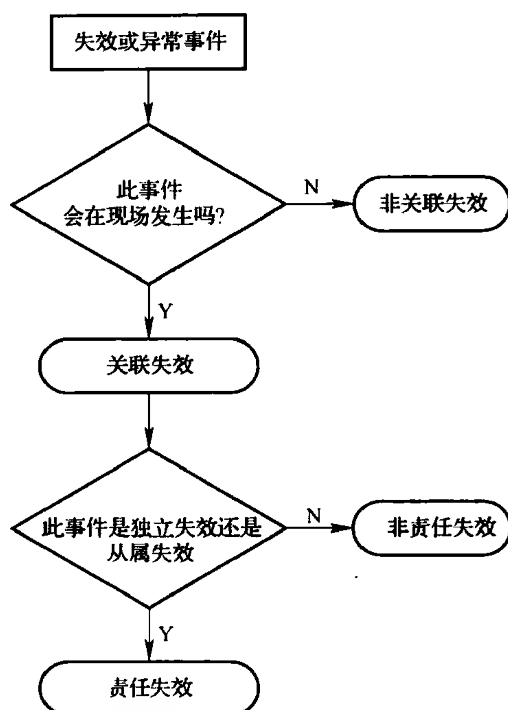


图 13.3 失效的分类流程

间歇性失效（Intermittent Failure）、不可重现失效以及操作错误引起的失效通常是有争议且难以分类的。这些类型的失效额外地增加了维修、保障和后勤人员的负担。这种类型的失效还令消费者感到沮丧，如某人的汽车经常间歇性熄火，但在修理时却无法重现这种故障；再如某人的电视机在家里每30min 会自动调整亮度一次，但在修理店却能够正常工作。典型的间歇性失效的原因是外部电源中断、电涌或瞬变电流等，它通常属于非责任失效。为了避免将间歇性失效分类为责任失效，在可靠性改进计划中必须使用外部电源监视器对输入电源进行监测、调控和记录。如果间歇性失效与外部电源中断、电涌或瞬变电流没有联系，那么就要把它作为“不可重现”（Cannot Duplicate，CND）失效进行调查分析。

CND 是随后的故障诊断和维修无法验证、核实或重现的试验事件。导致 CND 出现的常见原因是间歇性失效、机内测试不充分、操作错误、维护不当、管理误区和错误的用户手册等。在可靠性改进计划中，由于时间的限制，CND 往往容易被忽视，因为时间要用来尝试可重现失效，以验证这些设备是否处于良好的运行状况。为了维持试验的效率，以下的准则可以用来对 CND 进行分类：如果在 CND 的故障诊断中需要进行任一组件的调换、断开、重新连接以及位置调整，那么此类 CND 就可以归类为责任失效；如果在故障诊断中只使用了机内测试（Built-In Test，BIT）设备或只进行设备自我内部诊断（也就是说，对设备不进行任何的变更），那么就把此类 CND 归类为非责任失效。无论是以上哪一种情况，必须把所有的 CND 报告给产品故障报告、分析和纠正措施系统（FRACAS），并由已认证的测试性工程师或后勤工程师对其进行分析调查。在不进行可靠性改进计划的时期，工程师们可以尝试清除引发这些 CND 的原因。即使是那些

从可靠性观点上被归类为非责任失效的CND，往往也可能与维修性或测试性相关。在许多包括软件、使用机内测试的产品中，众多的CND和间歇性失效可能会导致软件代码中的漏洞或错误。考虑到可能存在的软件缺陷，应对这些失效进行评估。如果产品不包含软件或BIT，失效可能只在特定的环境条件下发生，如在某一湿度、温度或振动条件下发生。

虽然可靠性验证试验中的操作失误导致的失效总是属于责任失效，但可靠性增长过程中的此类失效的分类却值得商榷。在现场使用中，如果它们会危害到生命或造成其他灾难性的损失，那么就把它们归为责任失效；对于那些不会造成灾难性损失的失效，可以设立相应的分类准则，例如同样的操作失误发生了3次，就可以在第3次操作失误出现时将它归类为责任失效。操作人员不断重复的某些错误能够说明操作指导书或用户手册写得不好。例如在一个盒式磁带录像机（Video Cassette Recorder，VCR）的增长试验中，测试工程师需要每24小时对计时功能进行一次确认。测试工程师仔细遵循了VCR的操作说明，但他并不是每次都按要求打开计时功能。如果这种情况只发生一次，此失效可能会由于操作失误而被列为非责任失效；但如果它发生3次以上，则属于责任失效。

试验时间也可以分为关联试验时间或非关联试验时间。当设备处于正式试验状态，失效发生之间的试验时间称为关联试验时间。花费在设备失效的故障排除和核实维修是否有效的时间称为非关联试验时间。只有累积关联试验时间用来计算产品可靠性的改进。应当对发生在非关联试验时间内的失效进行分析和纠正，但不用它们来衡量可靠性。

13.2.3 试验优化

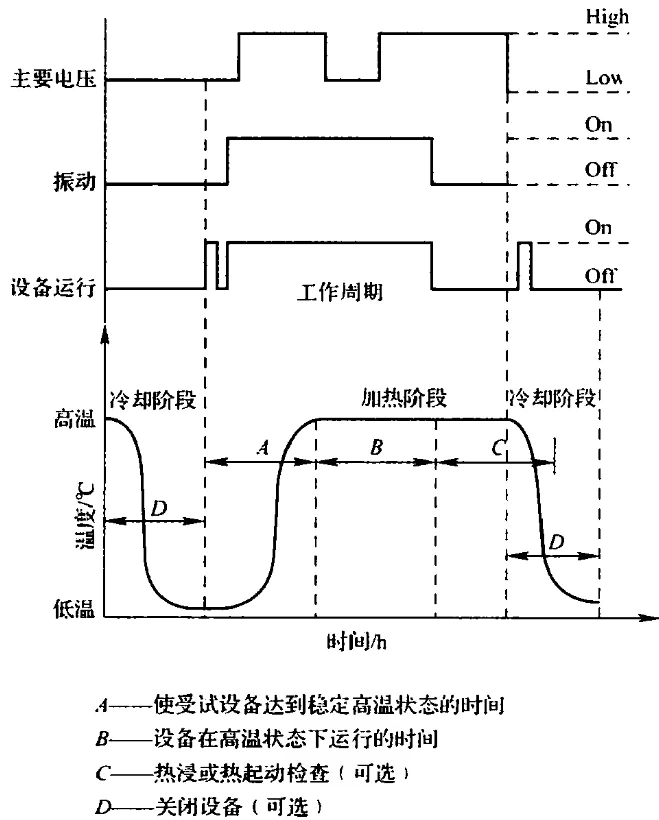
为了避免工作的重复，在进行增长试验的同时也应当进行其他类型的试验，如功能性、人为因素和安全试验。因为其他试验结果引起的设计变更可能会影响产品的可靠性，所以应该最大限度地共享各种试验得到的数据，这样才能更深入地了解设备的表现。BIT虚警核实（False-Alarm Verification）是可靠性增长试验中一个重要但往往容易被忽略的测试。许多产品都使用BIT来确定失效发生的时间。当BIT发现一个失效，但实际上这个失效并没有发生，这种情况被认为是虚警。通常用它占所有失效的百分比来表示虚警（通常在1%~5%之内），它属于CND的一个子集。若要为虚警验证使用增长试验，必须用外部测试和记录设备对产品的性能进行监测。在可靠性增长试验中，如果BIT和现场设备具有相同的触发敏感度，那么就可以用BIT数据和外部测试仪器得到的信息来计算虚警率。即使BIT的阈值不同，我们仍然可以用这些数据表示总的虚警表现。

13.2.4 试验周期和环境问题

产品在使用现场经历的相关试验时间包括一系列试验周期，这些周期结合了最坏的环境应力。第13.3节将讨论所施应力高于预期现场使用条件的加速试验。根据现场使用条件的不同，周期性试验的应力可能包括与电相关的（如输入功率的波动）、与热相关的、潮湿引起的（湿度）或振动引起的应力。一般情况下，可能不需要对最坏的环

境情况进行应力模拟试验，如家用消费类电子产品；但如果产品可能存在安全问题，就必须对产品进行最坏应力模拟试验。

为了促成失效的发生，试验中使用的环境条件往往是设备在现场使用中可能碰到的最坏的应力情况。图 13.4 是一个典型的周期性试验计划的例子——环境循环试验。实验室为设备提供快速的温度和电力变化，同时加以电动力或机械振动。对设备运行状况进行持续检查或在每个试验周期中定期进行检查，性能检查则不需要那么频繁。性能检查通常在室温条件下进行，它包括运行状况检查和额外的设备表现验证，如精度和重现次数等。应该在极端环境条件试验（如振动试验）期间或结束后立即进行性能测试，以进一步了解设备表现。



A——使受试设备达到稳定高温状态的时间
B——设备在高温状态下运行的时间
C——热浸或热起动检查（可选）
D——关闭设备（可选）

图 13.4　环境循环试验示例

13.3　应力余量试验

加速试验是一种可靠性改进技术，它通过增加产品的正常应力来迅速确定产品缺陷。加速试验的基本条件包括以下内容：

① 在正常应力和加速应力条件下产生的主要失效模式必须相同。

② 加速应力条件下，与失效机理相关的材料工程特性在试验前后应该相同。

③ 在额定应力和更高应力水平下，失效机理的失效概率密度函数的曲线形状应该相同。

要确定何时满足这些条件，必须确定失效模式（机理）。失效机理是导致失效的各种应力共同作用的过程，可能包括物理、电子、机械和化学等应力。失效模型中使用的这些应力则可以用来预测产品的可靠性。当上述的三个基本条件得到满足后，就可以使用加速寿命试验来减少试验的时间和成本。加速试验在产品的一般工作条件或者额定载荷下增加温度循环、振动、湿度和功率循环等应力。文献［Pecht，1991］研究出了基于温度、湿度、电压和机械应力的加速试验技术。我们可以根据在加速试验条件下得到的试验结果，推算出产品在正常运行条件下的等效失效时间。

如图 13.5 所示［Seusy，1987］，只有在应力超过强度时失效才会发生。产品的强度一般呈广泛分布，并会随时间的推移而减小，如图 13.6 所示。应力试验模拟产品的老化过程，并在这个过程中放大产品的不可靠度。图 13.7 显示了加速寿命试验背后的一般物理原理，我们将讨论应力寿命试验（Stressed Life Test，STRIFE）和高加速寿命试验（Highly Accelerated Life Test，HALT）的加速试验技术以及加速寿命试验模型［如幂律模型（Power Law Model）和 Miner 准则］［Schinner，1988；Hobbs，1990］。

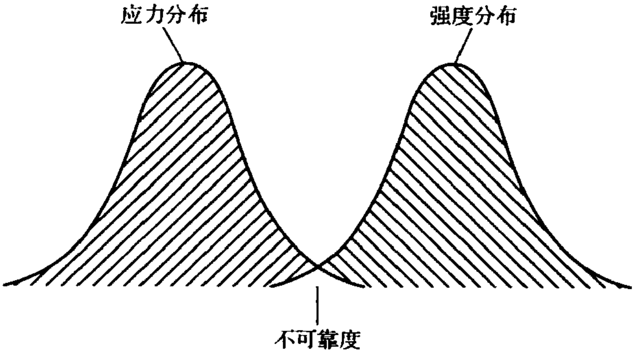


图 13.5 应力和强度

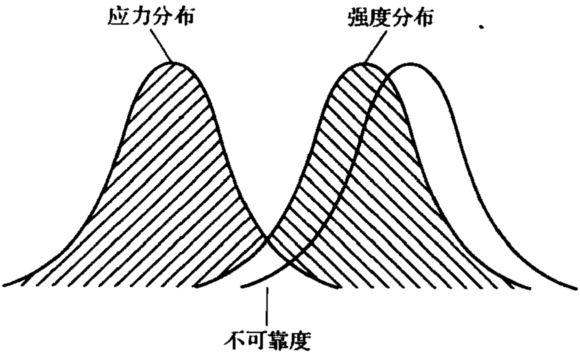


图 13.6 时间对强度的影响

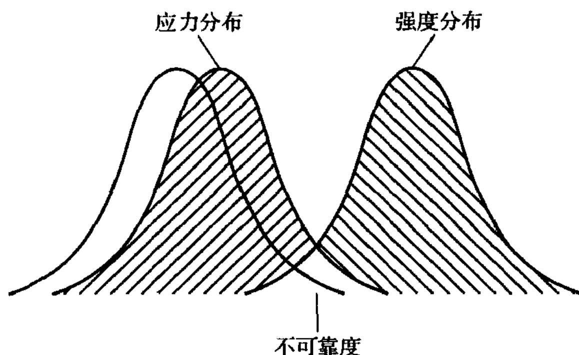


图 13.7 应力试验准则

13.3.1 应力寿命试验（STRIFE）

惠普公司采用的应力寿命试验（Stressed Life Test，STRIFE）使用温度循环、功率循环和（或）频率变换来加速产品的失效。从本质上讲，它与“正常”的可靠性改进方案相同，在测试过程中使用与产品实际运行相同的环境。通过增加温度变化范围、温度变化率和随机振动，惠普改进了 STRIFE，并将其应用到印制电路板上［电路板 STRIFE 试验（B. E. S. T.）］。进行 B. E. S. T. 的一些必要条件包括：

① 电路板上的元器件必须按照温度曲线维持在某一状态或者进行快速温变，温变范围是超出极端冷热温度 15℃的 90%。该温度曲线还需要根据产品和试验室进行调整，超温的持续时间应该以元器件至少达到 90% 的冷热极端温度为准。

② 产品必须进行开关机循环，以产生内部的温度循环，从而加快电子元器件的失效。当产品处于开机状态时，组件的温度会根据产品的功耗、发热量和热传导率升高。开关机循环也将诱发由电压和电流瞬时故障引起的电应力。

③ 把随机振动应用到两个轴向上，从而产生最大的机械应力，这样能确定是否会发生了过大位移。

13.3.2 高加速寿命试验（HALT）

高加速寿命试验（Highly Accelerated Life Test，HALT）由 Hobbs 工程公司提出，它为产品施加高于产品正常运行或非运行时应力水平的应力［Hobbs，1990］。常用的应力包括温度应力、振动应力和电压应力。HALT 使用步进应力方法，即逐步增加应力水平，直到达到产品的运行或者破坏极限。一旦发生失效，就要通过分析研究，对在这种应力下的产品进行设计补偿。首先为每一个应力重复这个过程，然后再在组合应力（如温度和振动的组合）下进行试验。只有达到了产品正常工作条件以上的设计安全余量限时（即产品的破坏限），才能结束 HALT。因此，HALT 的试验周期往往难以预测。最符合成本效益的做法是尽可能早地在产品的设计初期进行 HALT 试验。Hobbs 工程公司对 HALT 有以下总结：

① 一个通过激发失效来提高产品成熟度的信息收集工具和设计方法。

② 能快速找到失效机理。

③ 有助于在技术限制下提高产品可靠性。

④ 对于大多数公司，可能需要进行一些变化。HALT 可以应用于不同级别的生产过

程，从组件到最终产品。如果正确进行，HALT 可以提高产品可靠性，降低产品寿命周期总成本，进而提高客户满意度。

13.3.3　逆幂律模型和 Miner 法则

逆幂律模型（Inverse Power Law Model）和 Miner 法则可以用来计算加速寿命试验结果［Raheja，1990］。逆幂律模型可以将加速试验条件下的结果反推到正常的运行条件。该模型认为产品寿命与应力的 N_a 幂成反比。其中，N_a 是加速因子，它是 S-N 曲线的斜率：

$$\text{slope} = -1/N_a \tag{13.1}$$

逆幂律模型可以表达为

$$\left[\frac{\text{正常应力下的寿命值}}{\text{加速应力下的寿命值}}\right] = \left[\frac{\text{加速应力值}}{\text{正常应力值}}\right]^{N_a} \tag{13.2}$$

一旦加速试验完成，通过求解“正常压力下的寿命”，将得到正常工作条件下的等效测试时间。例如加速应力是正常应力的两倍，产品在加速应力下的寿命是 4h，N_a 等于 2，则在正常应力条件下的等效寿命就是 16h。要确定由此测试引发的累积损伤，就需要应用到 Miner 法则。

Miner 法则将累积损伤（Cumulative Damage）表示为

$$\text{CD} = \sum_{i=1}^{k} \frac{C_{Si}}{N_i} = 1 \tag{13.3}$$

式中，C_{Si} 为某给定平均应力 S 作用下的循环次数；在应力 S 的作用下，N_i 是将导致产品失效的循环次数；k 为应力个数。Miner 法则假设每个零件都有一个使用疲劳寿命，每一个应力循环都会消耗一定比例的产品寿命。当 CD 等于 1 时，累积损伤将引发一次失效。

13.4　对可靠性持续增长的监控

通过对发生在 TAAF 过程（或加速寿命试验）的失效进行设计更改，产品可靠性将会得到相应的提升。在可靠性增长试验过程中会得到一些失效数据，这些数据可用来估计现有产品改进率和可靠性的持续增长效果。

估计可靠性增长效果的目的是对项目进度进行合理的安排。很多时候，可靠性增长估计用来计算为达到某个可靠性水平所需的试验时间。在试验过程中的任意时间点对可靠性进行评估，可以确定是否如期改进了产品，是否合理分配了资源。已有的连续型和离散型可靠性模型都可用来估计可靠性的增长［Duane，1964；Lloyd and Lipow，1962］。

13.4.1　持续增长模型

持续可靠性增长模型是为可维修产品提出的，平均故障间隔时间（MTBF）可用来衡量可维修产品的可靠性。将 MTBF 和试验时间的关系图绘制出来，可以说明可靠性的增长。累计试验时间除以设备相应的累计失效数，就可以得到 MTBF 值。［Duane，1964］最早使用了这个概念。另一个连续型可靠性增长模型是 AMSAA（美国陆军装备系统分析中心）模型［Crow，1974］。

1. Duane 模型

Duane 在通用电气公司工作时发现：一些产品在试验过程中的总体改进趋势都与累

积失效率有关。这些产品包括流体机械设备、飞机发电机和喷气式飞机发动机 [Duane, 1964]。将所有产品的累计失效数和累计试验时间在重对数坐标上绘制出来，得到的拟合线几乎就是一条直线。该直线的斜率表示 *MTBF* 的增长率，在确定和修复设计缺陷过程中，它还表示可靠性改进计划的效力。在试验计划中，随着设计改进在产品中的应用，失效将会逐步减少。这种现象可以用式（13.4）表示：

$$\lambda_{\Sigma}=\frac{\sum F}{t}=Kt^{-\alpha_r} \tag{13.4}$$

式中 λ_{Σ}——累计失效率；

ΣF——累计失效次数；

t——累计试验时间；

K——产品初始失效率的常数；

α_r——增长率。

增长率 α_r 总是介于0和1之间，它模拟了递减失效率。增长率可能达到的最大值为1。在增长计划中，通常可以接受的增长率为0.4～0.5。

直到可靠性改进计划结束，预期得到的产品失效率都是瞬时失效率。当前失效率或瞬时失效率（λ_{I}）可以根据累计失效数（$\sum F$）计算得到：

$$\lambda_{\text{I}}=\lim_{\Delta\to 0}\frac{\Delta(\sum F)}{\Delta t}=\frac{\mathrm{d}(\sum F)}{\mathrm{d}t}=(1-\alpha)Kt^{-\alpha_r} \tag{13.5}$$

瞬时 MTBF 也可以用图估法得到：将失效数据绘制到重对数坐标上，*Y* 轴是 MTBF 的点估计值，*X* 轴是失效时间；再用一条直线拟合这些点，将这条直线向上平移 $1/(1-\alpha_r)$，就可以得到瞬时 MTBF 直线，通过它就可以估计瞬时 MTBF 了。

在可靠性增长计划中，Duane 模型也可用来绘制预期或计划增长曲线，以图形化表示可靠性增长的进度。绘制规划可靠性增长曲线的步骤如下：

① 确定可靠性目标。

② 基于同类产品的历史数据或者先期的试验数据，确定产品可靠性增长曲线的初始可靠性水平。

③ 初始化试验时间，使其等于设计更改开始对产品产生影响的时间 [Crow, 1986]。初始试验时间取决于到 t_i 时刻至少出现一个失效的概率。通过求解方程，1减去产品的可靠度方程等于63.2%（在 $t=$ MTBF 时，63.2%的受试产品都已失效）到95%的失效概率，t 即为初始试验时间；如果产品的可靠度函数服从指数分布，至少出现一个失效的概率为90%，那么失效概率的增加会导致预计总试验时间增加。

④ 基于产品的复杂程度、成熟度、所用技术、进行失效分析的努力和积极程度以及管理部门的支持程度确定增长率。

⑤ 绘制可靠性增长曲线，它是用来衡量可靠性增长的基准。

瞬时 MTBF 线与目标 MTBF 线的交点就是可靠性增长计划预期的试验时间。增长曲线只是评估项目进展的一个指引，它并不能保证可靠性目标一定能够达到。为获得可靠性的增长，并达到可靠性目标，必须查明产品的设计缺陷并执行纠正措施。

案例13.1

某可靠性增长试验计划将某航空电子系统的MTBF由目前的250h提高到1000h。初始试验时间（t_i）是250h，t_i是根据在这个时刻至少出现一次失效的概率为63.2%而估计得到的。如果一个同类产品在相同的增长试验中的增长率已达到0.3，求：

（a）累计MTBF达到1000h的试验时间。

（b）瞬时MTBF达到1000h的试验时间（假设试验可以运行的日历时间为6个月，并有充足的试验零部件，而且试验每天可以运行24h）。

（c）在重对数坐标上绘制出累积MTBF和瞬时MTBF直线。

解：（a）累计MTBF达到1000h所需的试验时间。

公式（13.4）是一个累积试验时间和累计MTBF表达式［General Electric Company, 1973］，使用它可以推导出累计试验时间t_c：

$$t_c = t_i\left[\frac{\theta_R}{\theta_i}\right]^{1/\alpha_r} \tag{13.6}$$

其中，t_i为初始试验时间，θ_i为初始MTBF，θ_R为累积或目标MTBF，α_r为增长率。因此，

$$t_c = 250\left[\frac{1000}{250}\right]^{1/0.3}\text{h} = 25398\text{h} \tag{13.7}$$

（b）瞬时MTBF达到1000h所需的试验时间。

利用公式（13.5），将初始失效率K转换成等效瞬时失效率，就可以推导出瞬时试验时间T：

$$\lambda_I = K(1-\alpha_r) \tag{13.8}$$

其中，λ_I是瞬时失效率，K是将被转换的失效率（即初始失效率）。因此，

$$\lambda_I = 0.004(1-0.3) = 0.0028 \tag{13.9}$$

使用前面步骤中方程的变化形式

$$T = 250\left|\frac{0.0028}{0.001}\right|^{1/0.3}\text{h} = 7735\text{h} \tag{13.10}$$

其中，T等于瞬时试验时间。

试验所需试验件：该试验运行6个月，每个月730h，所以总时间为4380h；如果试验效率为50%，那么有效试验时间为2190h；将方程式（13.10）得到的瞬时试验时间除以2190h，得到3.53，即需要4个试验件。

（c）累计MTBF和瞬时MTBF直线如图13.8所示。

2. AMSAA模型

AMSAA模型扩展了Duane制定的概念，它应用了Duane的累积失效率，但其基础也是Poisson分布。对AMSAA模型的解释如下：令$dm_1 < dm_2 < \cdots < dm_k$，其中，$dm_i$表示第$i$个设计更改（$dm$）发生时的累计试验时间。假设相邻两个设计更改之间的失效率是常数，如图13.9所示。令λ_i表示第i个设计更改时间段（$dm_i - dm_{i-1}$）内的失效率。基于恒定失效率的假设，在第i个时间段内的失效数N_i服从平均失效数为

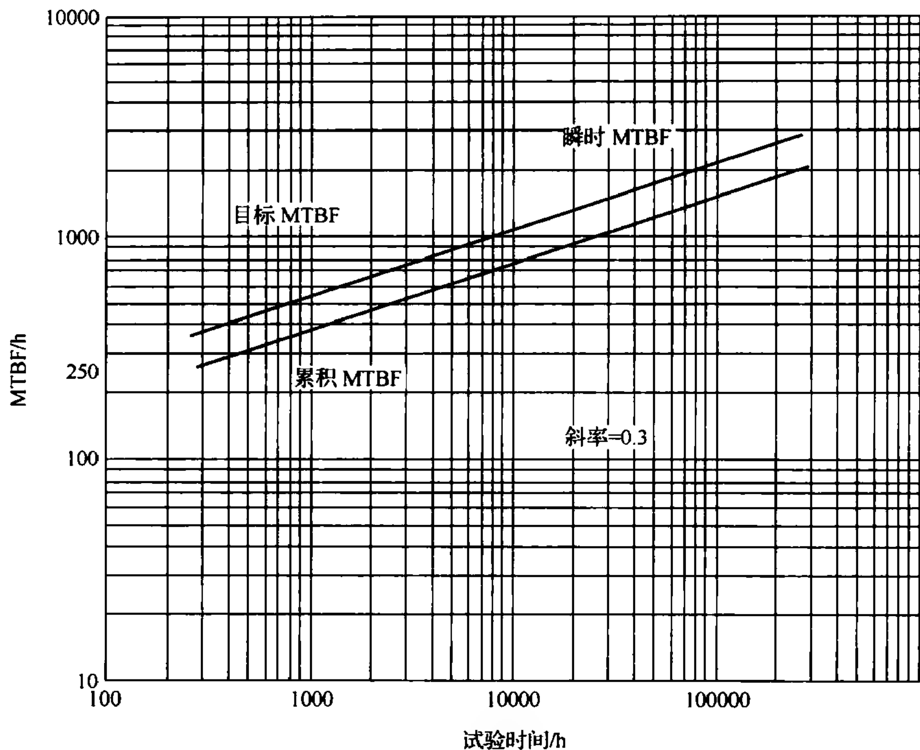


图 13.8 累积和瞬时 MTBF 直线

$\lambda_i(dm_i - dm_{i-1})$的 Poisson 分布，其数学表达式为

$$Prob[N_i = n] = \frac{[\lambda_i(dm_i - dm_{i-1})]^n e^{-\lambda_i(dm_i - dm_{i-1})}}{n!} \tag{13.11}$$

其中，n 是一个整数。

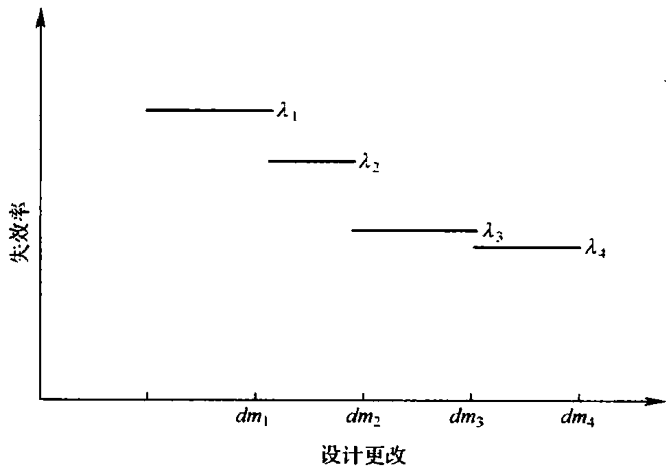


图 13.9 设计更改之间的恒定失效率

用t表示累积试验时间，$N(t)$表示到t时刻的总失效数。$N(t)$与Duane模型中的累积失效数ΣF类似。如果t在第一个时间区间内，则$N(t)$服从均值为$\lambda_1 t$的Poisson分布；如果t在第二个时间区间内，那么$N(t)$等于第一个时间区间内的失效数N_i加上第二个时间区间的失效数，即dm_1和t之间的失效数。因此，在第二个时间段内，$N(t)$服从均值为$\theta(t) = \lambda_1 N_1 + \lambda_2(t - dm_i)$的Poisson分布。

如果假定每个试验时间区间内（即设计更改之间）的失效率为常数（λ_0），则可以说，$N(t)$是一个均值为$\lambda_0 t$的齐次Poisson过程。如果失效率是随时间变化的，如图13.9中的从区间1到区间2，则可以说$N(t)$是一个非齐次Poisson过程。对于跟踪设计更改之间的可靠性增长，$N(t)$服从非齐次Poisson过程，其均值函数如下：

$$\theta(t) = \int_0^t \nu_\lambda(x)\,\mathrm{d}x \tag{13.12}$$

其中，强度函数$\nu_\lambda(x) = \lambda_i$，$dm_{i-1} < x < dm_1$。所以，对于任意$t$，有

$$Prob[N(t) = n] = \frac{[\theta(t)]^n \mathrm{e}^{-\theta(t)}}{n!} \tag{13.13}$$

其中，n为整数。当Δt趋近于零时，$\nu_\lambda(t)\Delta t$近似于在时间区间$(t, t+\Delta t)$内产品的失效概率。强度函数则近似于一个连续型参数方程，这样就可以编译试验数据并估计参数了。

AMSAA模型假定可以通过参数方程式（13.14）近似求得强度函数：

$$\nu_\lambda(t) = \lambda\beta t^{\beta-1} \tag{13.14}$$

式中的强度函数类似于公式（13.8）中的瞬时失效率λ_1。它也是Weibull失效率函数，$\beta>0$，$\lambda>0$，且$t>0$。由于AMSAA模型假设了一个包含Weibull失效率函数（不是Weibull分布）的Poisson过程，所以Weibull分布的统计过程在这里并不适用。公式（13.14）可以模拟包括可靠性增长在内的各种过程。因为$1-\beta$是可靠性增长率，所以参数β相对更重要一些。

通过参数估计，可得t时刻的平均失效次数为

$$\Theta(t) = \lambda t^\beta \tag{13.15}$$

这并不是产品的MTBF。在试验结束后，如果没有对产品进行额外的设计更改，那么产品将来的失效服从指数分布，产品的MTBF可以通过强度函数$\nu_\lambda(t)$的倒数求得。累积失效率$\omega(t)$为

$$\omega(t) = \frac{N(t)}{t} \tag{13.16}$$

这和公式（13.4）中的λ_Σ类似。如果重对数坐标上的累积失效率$\omega(t)$与时间有线性关系，那么这个可靠性增长过程就类似于Duane模型。

参数β和λ可以使用图估法在全对数坐标纸上求得或使用统计估计方法计算得到。图估法是在重对数坐标纸上将累积失效率（或平均失效数）与累积试验时间进行线性拟合。对累积失效率取对数，表示参数斜率和截距之间的关系。可以通过极大似然估计方法（已在第3章中有过讨论）对参数进行点估计。

在使用AMSAA模型之前，必须对试验数据加以分析，并找出明显的增长趋势而不

是齐次 Poisson 过程。可用来确定这种趋势的试验是中心极限定理试验或 Laplace 检验 [Cox and Lewis, 1966]。

在使用多个样机时，如 m 个样机，需基于累计试验周期（时间、里程等）对产品进行分析。类似于使用单台样机，对所有样机的失效数据进行综合分析。如果试验在某个失效发生时终止（即失效截尾试验），使用式（13.17）计算得到的检验统计量（μ_1）

$$\mu_1 = \frac{\sum_{i=1}^{m} X_i - mX_n/2}{X_n[m/12]^{0.5}} \tag{13.17}$$

其中，m 等于失效数（n）减 1，X_n 是最后一个失效发生的时间，X_i 是第 i 个失效发生的时间。如果是时间截尾试验得到的失效数据，使用式（13.18）计算得到的检验统计量（μ_2）

$$\mu_2 = \frac{\sum_{i=1}^{n} X_i - nt_0/2}{t_0[n/12]^{0.5}} \tag{13.18}$$

其中，n 是失效的数量，t_0 是总试验时间。对于指定的显著性水平 Z_α，统计量 μ 与标准正态偏差的关系如下：

① $\mu \leqslant -Z_\alpha$：在所选的显著性水平，可靠性增长明显，AMSAA 模型可用于估算参数 β 和 λ。

② $\mu \geqslant +Z_\alpha$：在所选的显著性水平，可靠性衰退明显，需要采取进一步的纠正措施或者进行设计改进。

③ $-Z_\alpha < \mu < +Z_\alpha$：在所选的显著性水平可靠性变化趋势不明显，因为数据（失效率）遵循齐次 Poisson 过程，需要积累更多的数据。

在正常分布表中，可以找到检验统计量的临界值。表 13.1 中列出了常用的双侧显著性检验统计值。

表 13.1　检验统计量

Z_α 值	显著性水平（双侧检验）
1.960	5.0
1.645	10.0
1.282	20.0

如果可靠性增长明显，可以通过公式（13.19）~式（13.26）求得 β 和 λ。有偏估计通常用于大样本。但如果通过了拟合优度检验，那么无偏估计也可用于小样本和大样本。对于失效截尾试验，β 的有偏估计值为

$$\hat{\beta} = \frac{n}{(n-1)\ln X_n - \sum_{i=1}^{n-1}\ln(X_i)} \tag{13.19}$$

β 的无偏估计值为有偏估计值乘以 $[(n-2)/n]$，即

$$\bar{\beta} = \frac{(n-2)}{(n-1)\ln X_n - \sum_{i=1}^{n-1}\ln X} \tag{13.20}$$

对于失效截尾试验（Failure Truncated Test），λ 的有偏估计值为

$$\hat{\lambda} = \frac{n}{X_n^{\hat{\beta}}} \tag{13.21}$$

λ 的无偏估计值为

$$\bar{\lambda} = \frac{n}{X_n^{\bar{\beta}}} \tag{13.22}$$

对于时间截尾试验（Time-Truncated Test），β 的有偏估计值为

$$\hat{\beta} = \frac{n}{n\ln t_0 - \sum_{i=1}^{n} \ln(X_i)} \tag{13.23}$$

β 的无偏估计值为有偏估计值乘以$[(n-2)/n]$：

$$\bar{\beta} = \frac{(n-1)}{n\ln t_0 - \sum_{i=1}^{n} \ln(X_i)} \tag{13.24}$$

λ 的有偏估计值为

$$\hat{\lambda} = \frac{n}{t_0^{\hat{\beta}}} \tag{13.25}$$

λ 的无偏估计值为

$$\bar{\lambda} = \frac{n}{t_0^{\bar{\beta}}} \tag{13.26}$$

拟合优度模型用来确定所收集的数据是否适合使用 AMSAA 模型。在第 12 章中讨论的 Cramer-Von Mises 检验是用来检验 AMSAA 模型的常用拟合优度检验法。根据选定的显著性水平 α，可以用表 13.2 确定检验统计量的临界值 C_m^2，然后将观测计算得到的结果与这个临界值进行比较。如果是失效截尾试验，这个临界值可以通过公式（12.72）计算得到：

$$C_m^2 = \frac{1}{12m} + \sum_{i=1}^{n}\left[\left(\frac{X_i}{X_n}\right)^{\bar{\beta}} - \frac{2i-1}{2m}\right]^2 \tag{13.27}$$

如果是时间截尾试验，C_m^2 可以通过式（13.28）计算：

$$C_m^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[\left(\frac{X_i}{t_0}\right)^{\bar{\beta}} - \frac{2i-1}{2n}\right]^2 \tag{13.28}$$

表 13.2　Cramer-Von Mises 统计的 C_m^2 参数形式的临界值

	重要性级别 α				
m	0.20	0.15	0.10	0.05	0.01
2	0.138	0.149	0.162	0.175	0.186
3	0.121	0.135	0.154	0.184	0.231
4	0.121	0.136	0.155	0.191	0.279
5	0.121	0.137	0.160	0.199	0.295
6	0.123	0.139	0.162	0.204	0.307
7	0.124	0.140	0.165	0.208	0.316

（续）

重要性级别 α					
8	0.124	0.141	0.165	0.210	0.319
9	0.125	0.142	0.167	0.212	0.323
10	0.125	0.142	0.167	0.212	0.324
15	0.126	0.144	0.169	0.215	0.327
20	0.128	0.146	0.172	0.217	0.333
30	0.128	0.146	0.172	0.218	0.333
60	0.128	0.147	0.173	0.221	0.333
100	0.129	0.147	0.173	0.221	0.336

如果计算结果大于表中所列的临界值，AMSAA 模型将不可用。Cramer-Von Mises 检验结果不理想，其原因是可靠性改进计划中的计划变更导致的跳跃或不连续。将数据绘制到坐标纸上，可以说明是否应该使用不同的模型或在何处发生了不连续。如果有跳跃发生，就可以在局部应用 AMSAA 模型；将跳跃或不连续发生前后的数据分开处理。如果计算结果小于表中所列的临界值，那么就可以应用 AMSAA 模型，产品强度函数［公式(13.14)］可以估计为时间的函数。一旦确定了强度函数，就可以通过绘制参数曲线预测出产品未来的可靠性表现。如果在时刻 t_0 后没有对产品进行修改，根据指数分布，失效率恒定为 $\lambda_0 = v_\lambda(t_0) = \lambda\beta t_0^{\beta-1}$。该 MTBF 的估计值就等于$[1/(\lambda\beta t^{\beta-1})]$。在AMSAA 模型的置信限表中可查得此 MTBF 的置信区间上下限。如果 $m>100$，使用$m=100$。

案例 13.2

某可靠性增长试验的累积试验时间是 2500h，产品在以下试验时间处发生了失效：第 85h、151h、184h、267h、378h、474h、660h、803h、1031h、1230h、1400h、1589h、1643h、1756h 和 2122h。在显著性水平为 10% 时，判断 AMSAA 模型是否适用？如果适用，确定 2500h 处的 *MTBF*。

$$\mu_2 = \frac{\sum_{i=1}^{n} X_i - nt_0/2}{t_0[n/12]^5} = -1.7806 \tag{13.29}$$

由于 -1.7806 小于 -1.645，表明在 10% 的显著性水平下可靠性增长明显，可以应用 AMSAA 模型估计参数 β 和 λ。β 的无偏估计是

$$\bar{\beta} = \frac{n-1}{n\ln t_0 - \sum_{i=1}^{n}\ln(X_i)} = 0.679 \tag{13.30}$$

λ 的无偏估计是

$$\bar{\lambda} = \frac{N}{t_0^{\bar{\beta}}} = 0.073942 \tag{13.31}$$

Cramer-Von Mises 统计量为

$$C_m^2 = \frac{1}{12m} + \sum_{i=1}^{m}\left[(X_i/t_0)^{\bar{\beta}} - \frac{2i-1}{2m}\right]^2 = 0.100305 \tag{13.32}$$

计算得到的临界值为0.100305，小于从Cramer-Von Mises表查到的值0.169。因此，可以应用AMSAA模型，在$t=2500\text{h}$处的强度函数为

$$v_{\lambda}(t)=\lambda\beta t^{\beta-1}=0.004074 \tag{13.33}$$

该强度函数的倒数就是求解的MTBF，即245h。

对这个例子中的数据进行修改，就可以说明通过AMSAA模型计算得到的MTBF结果对于失效时间的敏感性。如果前两次失效很早就出现在试验中——在第1h和3h而不是在第85h和151h发生，那么β将等于0.483094，λ将等于0.342426，在$t=2500\text{h}$处的MTBF为345h。为了确保失效不会在增长计划开始的时候发生，我们可以完成一些试验前的任务，如ESS、老炼、热测试等等。

文献［Crow，1988］研究了用于数据丢失或者某些失效时间未知的可靠性增长估计技术，这种技术在处理某些试验间隔时间内的情况时非常有效。总体来讲，分组数据的估计方法较为复杂，因为不存在可用于计算β的闭型方程。假设有k个时间间隔，间隔范围为$x=x_0$，x_1，…，x_k，那么，β可以通过方程式（13.34）估计得到：

$$\sum_{i=1}^{n} n_i \frac{x_i^{\hat{\beta}}\ln x_i - x_{i-1}^{\hat{\beta}}\ln x_{i-1}}{x_i^{\hat{\beta}} - x_{i-1}^{\hat{\beta}}} - \ln x_k = 0 \tag{13.34}$$

式中，$x_0\ln x_0$定义为零。必须采用一些数值计算方法来对β进行求解。如果估计得到了λ后，$\hat{\lambda}$可以通过式（13.35）估计得到

$$\hat{\lambda}=\frac{\sum_{i=1}^{k} n_i}{x_k^{\hat{\beta}}} \tag{13.35}$$

还有一些其他连续型可靠性增长模型，这些模型在适用于某些研发项目或特定条件的前提下，可用来对产品可靠性进行建模。其中，包括用于硬件可靠性增长的Cox-Lewis模型（1966）和Lloyd-Lipow模型（1962）等；用于软件可靠性增长的模型，如Jelinski-Moranda模型（1972）和Littlewood-Verrall模型（1973）等。

13.4.2 离散模型

Duane和AMSAA可靠性增长模型都是连续型模型，它们适用于以时间段来衡量可修复产品的可靠性。离散型模型与连续型模型不同，因为用它们来衡量可靠性的产品是运行或停机的状态，如导弹或火箭。对于服役中的产品，可以使用离散型函数表示其失效或者运行的状态。常用的离散型模型有Lloyd and Lipow模型（1962）和Wolman模型（1963）。

1. Lloyd-Lipow模型

Lloyd-Lipow（1962）有两个模型：第一个模型假设产品在可靠性改进计划中只有一种失效模式。如果其失效模式尚未消除，那么每次试验都假设该产品发生失效的概率是常数；如果前一个试验成功结束，且产品没有失效，就进行下一个试验；如果产品出现失效，则尝试执行纠正措施或变更设计来消除这个失效模式，此失效模式的消除概率也被假定为常数。因此，在进行第n个试验时，产品的可靠性R_n为

$$R_n=1-Ae^{-C(n-1)} \tag{13.36}$$

其中，A 和 C 是预先设定参数。

Lloyd-Lipow 的第二个模型将改进计划分 k 个阶段进行。在第 i 个阶段，对 n 个产品进行试验。记录每个阶段的结果和发生的失效，直到该阶段试验结束，才对产品可靠性进行改进。此时的可靠性增长函数为

$$R_i = R_\infty - [\alpha_r / i] \tag{13.37}$$

其中，R_i 是第 i 个阶段的产品可靠性，R_∞ 是当 $i \to \infty$ 时的可靠性极限值，α_r 是增长率（$\alpha_r > 0$）。极大似然估计和最小二乘估计可以用来确定 R_∞ 和 α_r。最终或是第 k 阶段的可靠性可以通过置信下限来确定。

2. Wolman 模型

Wolman 模型（1963）认为产品发生失效的状况是由附带原因（Inherent Cause）或由于某个可指明原因（Assignable Cause）造成的。对于每一个试验——如导弹发射——其结果可能是成功或失败。如果失败了，就需要确定其原因是附带的还是可指明的。附带原因导致的失效反映了产品的制造工艺，无法通过纠正措施得到消除；可指明原因导致的失效可以得到消除。Wolman 模型假设：可指明原因导致的失效模式是已知的，当这些失效模式中的一种导致产品失效，它将被永久地从产品中消除。Markov 链（Markov-Chain）方法可以用来确定经过 n 次试验后的产品的可靠性。该模型的表达式为

$$R(n) = \sum_{i=0}^{k} (1 - q_i)(1 - q_0)^{k-i} P_{0,i}^{(n)} \tag{13.38}$$

其中，q_i 是固有失效模式导致产品失效的概率；q_0 是由可指明原因导致的产品失效概率；$P_{0,i}^{(n)}$ 是第 n 步时的跃迁概率。除了 Lloyd-Lipow 模型和 Wolman 模型，还有一些其他可用的离散型可靠性增长模型，如 Barlow-Scheuer 模型（1966）和 Singpurwalla 模型（1978）。

13.5 可靠性改进的效率和不确定性

所有可靠性改进计划的总体目标都需要确定、纠正并消除设计和制造的缺陷及失效模式。如果可靠性改进计划是按计划进行的，设计缺陷就能在增长试验而不是使用现场中暴露出来并得到纠正。但两个现象约束了这一过程：即可靠性增长试验的有效性和不确定性。

13.5.1 可靠性增长效率

可靠性增长试验中出现的失效可能会得到纠正，也可能不会。不能得到纠正的失效模式为 A 类失效，得到纠正的失效模式为 B 类失效，文献［Wolman，1963］也曾讨论过类似的失效分类方法。在对 B 类失效模式的可靠性数据进行分析后，需要仔细地实施纠正措施。在理想情况下，所有关联失效模式都应予以纠正，然而由于资金或目前的设计工艺水平的限制，会存在一些 A 类失效。经验表明，平均有 30% 的 B 类失效模式仍继续存在于产品中，即使它们被认为已经得到了纠正。将被消除的 B 类失效的比例等于可靠性增长效率因子。试验完成后的潜在增长效率可以确定为

$$System_{GP} = 1/[\lambda_A + [(1 - EF)\lambda_B]] \tag{13.39}$$

其中，$System_{GP}$是产品潜在增长效率；λ_A 是观测到得 A 类失效模式的失效率；λ_B 是观察得到的 B 类失效模式的失效率；EF 是可靠性增长效率因子。

案例 13.3

某可靠性增长试验在3000h 处结束，发现25 个失效。通过失效分析，确定有6 个 A 类失效，19 个 B 类失效。同类产品的经验表明，效率因子应该是 70%。此产品的潜在增长效率为

$$System_{GP} = \left\{1/\left[\frac{6}{3000} + \left[(1-0.7)\times\frac{19}{3000}\right]\right]\right\}\text{h} = 256\text{h} \tag{13.40}$$

这个例子说明了在试验完成后，如何使用已知的累积试验数据确定潜在的增长效率。在可靠性增长试验开始之前，Duane 模型也可使用类似的分析方法，根据不同的增长率，确定为了达到规定的 MTBF 所需的试验时间。

13.5.2 可靠性增长的不确定性

国防部进行了一项关于 MTBF 和可靠性增长估计的不确定性的研究［U.S. Air Force, Army, and Navy, 1989］。在研究初期，一个 AMSAA 模型的实例显示：发生在试验初期的失效会造成40% 的 MTBF 增幅（245 ~ 345h）。通过可用于增长试验数据的分析、失效的分类、增长试验的效率的大量模型，可以看出造成分析结果有巨大差异的原因。美国空军、陆军和海军使用 AMSAA 模型进行蒙特卡罗（Monte Carlo）模拟，以确定 MTBF 和可靠性增长估计值的不确定性。对于每个蒙特卡罗试验，记录所有前 30 次失效，并对增长率和瞬时（或当前）的 MTBF 进行估计（并在图上标注出来）。图 13.10 所示的数据范围包含了 80% 的仿真结果。在 5 次失效之后，10% 的 MTBF 估计值将可能大于真实值的 2.6 倍，10% 将低于 0.45 倍的真实值。图 13.11 中的三个图表也是利用蒙特卡罗模拟得到的。这些图表说明，随着增长率真实值的增加，增长率估计值的分散程度会随之削弱。

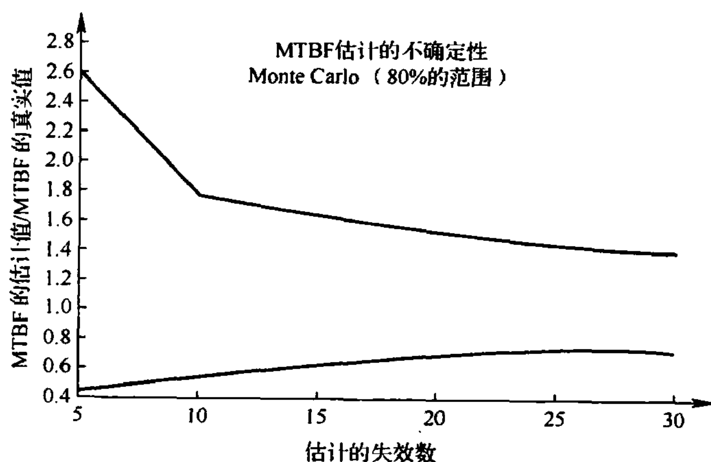


图 13.10 MTBF 估计的不确定性

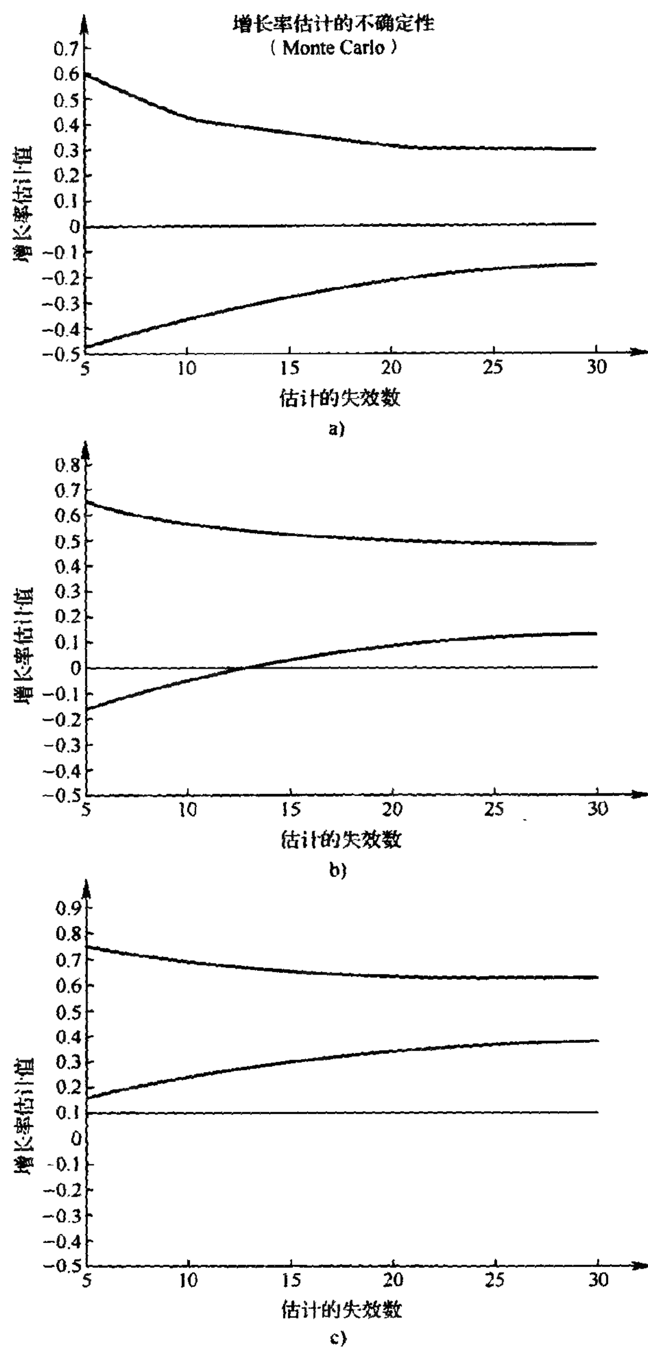


图 13.11　MTBF 估计的不确定性

a）真实增长率 =0.1　b）真实增长率 =0.3　c）真实增长率 =0.5

这些 Monte Carlo 模拟所展现出的差异以及在增长试验的规划和实施中固有的差异使我们确信：不能仅依据增长试验的结果做出关键计划的决策。但这并不意味着可靠性增长试验不具有经济效益。事实上，这是持续提高产品可靠性的一个高性价比方法。这也意味着，应该用良好的工程决策和其他研发试验或分析的结果一起比较增长计划的结果，如热相关试验和可靠性预测等。此外，增长试验所得到的可靠性结果应该由某个置信区间的上限和下限进行约束。在此情况下，我们可以说 MTBF 的真实值将处于某个置信区间的上、下限之间；如置信度为 80% 时，MTBF 的真实值会有 1/5 的机会处于该范围之外。

如果可靠性增长计划的实施目的是为了消除产品缺陷，那么产品的可靠性将在这个过程中得到提升，并有可能不需要进行可靠性验证试验；但如果对增长试验的结果或主要计划决策的必要性有疑问时，那么就可以在必要的时候考虑进行可靠性验证或鉴定试验。

13.6 总结

本章讨论了可用作持续改善产品可靠性的技术。可靠性改进过程认为一个处于设计阶段的复杂产品的可靠性可以得到改善，并可以为其分配改进时间。在一定条件下，对产品进行运行或试验，可以确定由于设计、制造和（或）运行造成的缺陷，检测并消除这些缺陷。可靠性设计方法在这个过程中可用于重新评估或改进产品可靠性。本章特别讨论了可靠性增长过程的原理、强度余量试验、持续可靠性增长监控方法、可靠性改进的有效性和不确定性。

关于可靠性增长过程的内容详细阐述了包括实施可靠性增长计划的必要性、失效分析流程概述、常见失效模式以及确保可靠性增长计划成功的失效分类技术等。关于强度余量试验的内容包括加速试验技术及衡量加速试验效果的工具。关于持续可靠性监控的内容讲解了连续型和离散型增长模型。我们还对常用的 Duane 和 AMSAA 模型进行了详细讨论并举例说明。最后一节提出了判断纠正措施有效性和可靠性增长技术不确定性的技术，并给出了相关案例。整章内容有助于技术人员和管理人员以高成本效益的方式不断提高产品的可靠性。

参考文献

Barlow, R. E. , and E. M. Scheuer. 1966. Reliability growth during a development testing program. Technometrics 8: 53.

Cox, D. R. , and P. A. W. Lewis. 1966. The statistical analysis of series of events. New York: John Wiley & Sons.

Crow, L. H. 1974. Reliability analysis for complex repairable systems. Technical report # 138.

U. S. Army material systems analysis activity, Aberdeen Proving Ground, Aberdeen, MD.

———. 1986. On the initial system reliability. Proceedings of the Annual Reliability and Maintainability Symposium, Las Vegas.

———. 1988. Reliability growth estimation with missing data—II. Proceedings of the Annual Reliability and Maintainability Symposium, Los Angeles.

Duane, J. T. 1964. Learning curve approach to reliability monitoring. IEEE Transactions on Aerospace 2 (2): 563.

General Electric Company. 1973. Research study of radar reliability and its impact on life-cycle costs for the APQ - 113, - 119, - 120, - 144 radars, Utica, NY.

Hobbs, G. K. 1990. Highly accelerated life tests—HALT. Westminster, CO: Hobbs Engineering Corporation.

Jelinski, Z., and P. B. Moranda. 1972. Software reliability research. In Statistical computer performance evaluation, ed. W. Freiberger. New York: Academic Press.

LeStrange, J. 1990. Failure analysis laboratory. Litton Amecon briefing to the University of Maryland Reliability Engineering Program.

Littlewood, B., and J. L. Verrall. 1973. A Bayesian reliability growth model for computer software. Record IEEE Symposium on Computer Software Reliability, New York.

Lloyd, D. K., and M. Lipow. 1962. Reliability: Management methods and mathematics. Englewood Cliffs, NJ: Prentice Hall.

Pecht, M. 1991. Handbook of electronic package design. New York: Marcel Dekker.

Raheja, D. G. 1990. Assurance technologies: Principles and practices. New York: McGraw-Hill.

Schinner, C. 1988. The board electronic STRIFE test (B. E. S. T.) program. Reliability Review 8: 3.

Seusy, C. J. 1987. Achieving phenomenal reliability growth. Proceedings of the ASM Conference on Reliability—Key to Industrial Success, Los Angeles, CA, 1987.

Singpurwalla, N. 1978. Estimating reliability growth (or deterioration) using time series analysis. MIL - HDBK - 189, appendix D.

U. S. Air Force, Army, and Navy. 1989. The TAAF process, appendix C—Uncertainty of MTBF and growth estimates. HQ AMC/QA, OASN S&L, HQ USAF/LE - RD, C1-C3.

Wolman, W. 1963. Problems in system reliability analysis. In Statistical theory in reliability. Madison: University of Wisconsin Press.

第 14 章　后勤保障

14.1　引言

所有用户都希望他们所购买的产品不会发生故障，也不希望产品在预期使用寿命内出现故障。此外，如果产品出现了故障或性能降低的情况，一般来说，用户都期望能对产品进行及时的维修或更换，但相对于产品购买价格来说，维修或更换的费用一定要合理。另外，用户希望产品只需要很少或者根本不需要预防性维修或定期维修，因为这样可以将产品的使用成本和不可用性都降到最低。

在民用产品领域，通过消除或最小化预防性维修（Preventive Maintenance，PM），工业行业已经提高了产品的可靠性和耐久度。在汽车行业，减少预防性维修的例子包括电子设备、无触点点火产品以及长寿命润滑（Lubricated-For-Life）轴承的应用。很多产品零部件（如前轮轴承、等角速万向节组件和排气装置）的使用寿命都得到了延长，以至于用户在拥有产品的阶段不再需要更换这些昂贵的零部件。最终，保养和维修还是必要的，并且某些类型的维修必须由汽车经销商或独立销售商提供。用户更倾向于等待维护保养的完成；如果这样不可行的话，那么汽车也应该在当天，至少也要在第二天内完成维护保养。为实现这样的目标，经销商或修理店需要具备经过培训的人员、合理的设备、工具、检测设备、技术数据以及备用零件。这些后勤资源必须预先就位，以实现快速周转。因此，制造商必须在产品设计和研发期间投入时间和资金。影响产品维修或更换的必要的计划、采办和资源配置称为后勤保障（Logistic Support）。

为了满足用户的需求和期望，从生命周期的角度进行产品研发是非常有必要的。为了使产品能有效地满足用户在成本效益方面的基本需求，在产品设计和研发的初始阶段，设计师和项目管理人员必须考虑产品的可靠性、维修性以及效能。然而，在没有后勤和产品保障能力，或者不能根据产品应用情况将两者有机结合的情况下，用户的需求得不到完全满足，所以不仅要定义主要设备和应用软件的可靠性、维修性以及效能需求，还要定义包括后勤保障要素的资源需求方面的可靠性、维修性以及效能需求。

应用于产品的综合后勤保障（Integrated Logistic Support，ILS）包括对产品进行生命周期维修和保障的方法。综合后勤保障是产品规划、设计、研发、检测与评估、生产制造、应用和报废等环节中不可或缺的组成部分。后勤保障中，与研发人员相关的内容包括技术维护计划、供应保障、产品支持、包装、装卸、储存、运输、人力及人事、培训及培训支持、设备、技术数据、计算机资源支持和设计接口等。Blanchard（1992）提出了关于生命周期中的后勤保障的全面概述。本章将讨论可靠性对后勤保障需求的影响，着重分析产品、设备或组件的可靠性对于备件或维修配件、保障设备以及维修人员

需求的影响。

14.2 后勤保障要素

后勤保障包括在产品预定生命周期中，为产品提供经济有效的保障所必需的规划及管理、设计研发、资源的采办及配置。为了保证同时满足性能需求和成本需求，必须将后勤保障要素与产品的其他部分加以整合。后勤保障主要包括以下几个要素：

1）技术维修计划（Maintenance Plan）包括在生命周期中对产品全部保障的所有规划和分析。此过程由后勤保障分析（Logistics Support Analysis，LSA）、维修级别分析（Repair Level Analysis，RLA）以及后勤保障分析中的记录文档组成。

后勤保障分析（LSA）是一个迭代过程，它取决于来自可靠性和维修性预测、失效模式、影响及危害度分析等方面的输入内容。这些分析结果，加上设计评审和审查都由后勤工程师来完成，如果维修级别分析（RLA）认定产品是可更换的且（或）可维修的，那么LSA用来为每个产品、子组件及组件制订维修方案，以确定支持产品所必需的后勤资源。

与生命周期成本分析类似，维修级别分析也需要考虑在生命周期内支持产品所需的全部成本。维修级别分析是一种在特定维修级别上对产品及其组成部分进行维修，或报废成本效益的经济评价。维修级别一般分为三种：基层级或现场级，即用户级（级别O），在这个级别上，通常只能进行组件的拆除和更换；中继级（级别I），在这个级别上，可以利用现场或非现场（外场）的维修设备对组件和子组件进行维修；基地级（级别D）或制造商级，此级别可以返修或重新生产子组件。在上述的每一个维修级别上，维修级别分析都用来核算与维修或报废备选产品相关的成本。维修级别分析的结果会反馈给后勤保障分析（LSA），这样，最后的后勤保障分析就能在生命周期中反映出所需的维修方案和后勤资源，它们是开发或支持产品必不可少的。

2）供应保障（Supply Support）包括所有的备件（例如单元、组件、模块）、维修配件、消耗品、特殊补给品以及相关的库存，它们为主要面向应用的产品、软件、产品的测试和保障、产品的运输和装卸、培训设备以及其他设施提供支持。供应保障也包含物资供应文档、采购职能、仓储以及物料的配送，除此之外还包括在所有适用场合下，与备件及维修配件库存的购置和维护相关的人员。需要考虑的事项包括：每个维护等级、每个备件、维修配件配送和存储的地理位置、备件的需量率以及库存等级、库存点之间距离、采办期以及物料的配送方式。

3）保障及试验设备（Support and Test Equipment，STE）包括所有工具、特定条件监控设备、诊断及检验设备、计量及校准设备、维护站以及为产品的计划维修和非计划维修提供保障的维护保养和产品装卸。必须在每个维护级别上确定试验及产品保障需求，除此之外，还要确定对于基本标准或二次标准可追踪试验的所有需求。试验及产品保障可分为特有型（Peculiar）（对于新设计产品或用户库存中独有的成品以及开发中产品的试验）和普通型（Common）（对用户库存中的现有产品进行试验），也可分为专用

型（Special Purpose）（为支持研发中产品而特别设计的）和通用型（General Purpose）（一般是对成品进行试验，以保障最终产品，另外还包括对开发中不需要修改的产品的试验）。

4）包装（Packaging）、装卸（Handling）、存储（Storage）及运输（Transportation）包括所有特殊规定、集装箱（可重复使用的和用完即丢弃的）以及为保障主要产品的包装、保存、存储、装卸或运输的供给、测试及产品支持、备件及修理配件、人员、技术资料和移动设备。此要素包括产品初始配送以及以维修为目的的人员和物料的运送。

5）人力（Manpower）及人事（Personnel）包括产品的安装、检测、操作、装卸、持续维修以及相关的测试和产品支持所需的人员。人员的需求通过数量和技术级别来定义，它是根据保障级别和地理位置，为每个操作及维修功能而定义的。

培训及培训支持包括初期培训，其目的是使人员熟悉产品，除此之外，还包括为补偿替代人员的成长和消耗而提供的补充培训。培训能使受训人员的技术级别提升到产品所需要的水平。培训支持还包括那些为人员培训工作提供保障的辅助设备（例如模拟装置、实物模型、特种产品、软件等）。

6）设施（Facility）是指在各个级别上，所有产品运行和实现维修功能所需要的物理单元：物资设备、场地、简易建筑、房屋、中继级修理车间、校准实验室、特殊仓库修理和大修设备。资本设备和公用工程（如热能、电力、能源需求、环境控制、通信联络）通常也是设施的一部分。

7）技术数据（Technical Data）包括安装和校验工序、操作和维修说明、检验和校准程序、大修程序、修正说明、设施信息、图样以及必要的说明书，它们用来保证产品的运行和维修。这些数据不仅涵盖主要的应用设备，也包括测试和保障设备、运输和装卸设备、培训设备以及其他设施。

8）计算机资源支持（Computer Resources Support）包括在各个级别上对产品进行维修的所有计算机设备及辅助装置、状态监测及维修诊断辅助设备、软件、程序带、磁盘以及数据库。

9）设计接口（Design Interface）将后勤设计参数与产品准备、资源需求以及保障成本相关联起来。这些设计参数包括产品可用性或是所需求产量的实现情况、遵守地方性和全国性的环境及安全的法律法规、减少对能源的使用、所设计产品的应用能力或在使用中将其变成多于一个最终产品的容易程度。

14.3 可靠性对后勤资源的影响

以后勤专家的观点来看，可靠性可转化为对于后勤资源的需求，维修性可转化为保障主要产品运行所必需的后勤资源范围以及具体后勤资源（例如人员和产品支持）专注于单一的维修动作的持续时间。可靠性和维修性相互作用的结果是对后勤资产的需求，它们用来在用户期望的时间内，将运行准备状态或可用性维持在某一级别。运行可用度的公式为

$$A_0 = \frac{\text{MTBM}^{\ominus}}{\text{MTBM} + \text{MDT}} \tag{14.1}$$

在这一章中，我们将通过检验平均维修时间（Mean Time To Repair，MTTR）和影响它的保障反应时间来探讨平均故障间隔时间（Mean Downtime，MDT），还将检验可靠性对供应保障、物资供应以及产品支持和维修人员的应用而产生的影响。

14.3.1 可靠性、维修率及后勤资源的预期需求

低可靠性会导致后勤资源需求的增加。到达率是一个产品或者产品总体对后勤保障的要求。对于一个可靠性由齐次 Poisson 过程（也就是恒定故障率）表示的产品来说，其总需量率（Demand Rate）的计算公式为

$$\text{需量率} = \text{使用中的单元数量} \times \text{单位时间内每个单元的维修行为} \tag{14.2}$$

每个单位时间内维修行为的数量是产品关联、责任失效的函数，除此之外，它还是由于不完善的机内测试（Built-In Testing）或产品监测（虚警和虚假故障指示）引起的拆装的函数；由不完善的诊断程序导致的不必要的功能单元拆除的函数；或是为了拆除另一个组件而完成的拆除（也就是非关联、非责任失效）的函数。图 14.1 阐明了这些事件的潜在影响。

虚警（False Alarm）通常是一个由机内测试产品（BITE）发出的指示，它表明监测产品有错误发生，即便操作者没有发现任何性能恶化。如果操作者报告了一个问题，维修人员就可能会尝试重现此问题，这可能会触发一个无法重现失效（CND）或无失效发生（NFF）。然而，在后勤专家看来，此过程已经消耗了后勤资源。

如图 14.1 所示，如果操作者发现了一个故障，并向技术人员（O 级）请求援助，可能发生两个事件中的一个，或许维修人员能够重复故障条件，并为某“大概”失效的可替换或可维修产品（Repairable Product，RI）诊断出需要隔离的故障；或许维修人员并不能发现或重复此故障条件。

一项维修工作会产生多个结果。维修人员完成维修，或拆除、替换并执行检测程序，其结果可能是维修成功，恢复产品的全部功能；也可能是原来故障条件的重现（如果更换了错误的部件）。如果用以影响维修的零部件本身就是可维修的，那么失效的零件将交由更高级的维修设施维修（I 级或 D 级），也有可能返回到生产厂商进行维修。

如果更换两个或两个以上的可维修零部件，那么很有可能至少有一个零部件处于正常工作秩序（一个 RETOK 事件），对于电子设备来说尤其如此。在某些情况下，一个产品交由修理厂时，可能会发现它仍能运转，并且能满足最低性能需求，但是维修活动可能会要求对产品进行整修恢复，以延长其有效使用寿命。为了满足 I 级或者 D 级维修的需求，后勤保障必须再次提供所需的资源。

如果 O 级维修人员更换了一个或更多部件，而产品依然存在故障，后勤保障仍然必须提供同样的资源，这与真正发生了故障一样。如果更换了可维修组件，但所维修产

⊖ MTBM（Mean Time Between Maintenance）为平均维修间隔时间。

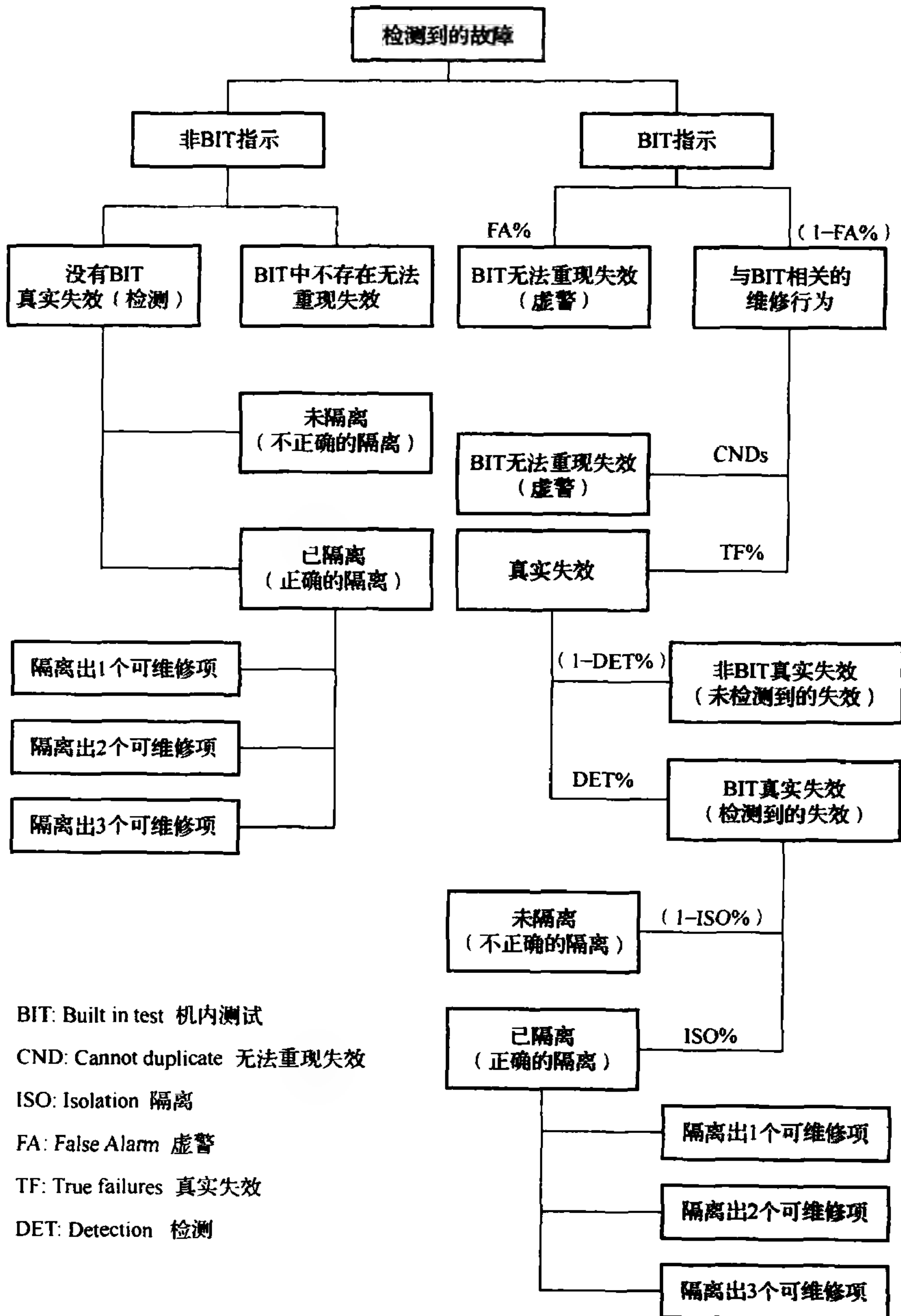


图 14.1 FA、CND 以及故障隔离与可替换项维修之间的关系

品仍然不能恢复原有的功能，那么 O 级——也可能在 I 级和 D 级维修上浪费了成本（时间和金钱）。故障的解决可能需要使用补充诊断程序或经验更丰富的 O 级维修人员，需要 I 级或 D 级的技术人员或制造商现场代表在现场或远程帮助，或是将产品送至更高级的维修设施。

因为在定义虚警（FA）、无法重现失效（CND）、重复测试通过（RETOK）的组成事件时所存在的差异，此处只讨论 FA 和 CND。首先，FA 是由 BITE（机内测试设备）指示，或是由一些其他形式的远程监控设备指示的，它表明产品出现了问题，即便操作者并没有发现；其次，无法重现（CND）率包括所有非运行维修行为和车间维修行为，在此期间，无法确定故障。

1. 虚警率

对于给定产品来说，FAR（False Alarm Rate，虚警率）的定义是 FA（虚警）次数除以用户（操作者）级别故障指示的总量。一般来说，对于包含 BITE 或其他自动、半自动监测设备的系统、子系统来说，只考虑用户级别的 FA 事故。FAR 可以用百分数或小数来表示，即

$$\mathrm{FAR} = \frac{\text{虚警次数}}{\text{虚警次数} + \text{真实故障数}} = \frac{\text{虚警次数}}{\text{用户故障指示总数}} \tag{14.3}$$

另外，

$$1 - \mathrm{FAR} = \frac{\text{真实故障数}}{\text{用户故障指示总数}} \tag{14.4}$$

如果 FA 事件不会引起维修请求，它就不会对后勤保障产生影响。

2. 无法重现率

在产品体系的任意一个级别（也就是系统、子系统、部件、装配组件、模块）都会发生无法重现（Cannot Duplicate，CND）事件，但是只能由维修人员触发。对于一个产品来说，给定维修级别上的 CND 率的计算公式为

$$\mathrm{CND} = \frac{\text{未发现故障的维修行为数}}{\text{未发现故障的维修行为数} + \text{真实故障数}} = \frac{\text{未发现故障的维修行为数}}{\text{维修行为总数}} \tag{14.5}$$

和

$$1 - \mathrm{CND} = \frac{\text{检测到的真实故障数}}{\text{维修行为总数}} \tag{14.6}$$

3. 故障检测概率

用户和后勤人员共同关注的是机内测试产品的失效，或是用户所要检测的故障。当操作者和 BITE 都不能发现所有的故障时，产品后勤保障会受到怎样的影响？应该关注后勤人员吗？如果由潜在过应力引起的失效，或部件中其他组件的后续失效所产生的当前和未来影响完全是良性的，那么后勤保障系统完全不会为此类型失效制定维修计划，或对其进行维修。如果检测出潜在失效产品，并进行了隔离和更换，那么此过程很有可能是在显性故障的维修工作期间完成的。潜在失效产品的检测和隔离是中继级或基地级保障产品更严格公差限的结果。

BITE 未发现的失效模式会产生不良影响，产品后勤保障必须提供一个备选的检测

和修复方法。这意味着，后勤人员的规划必须为操作者提供检测产品非BITE失效模式的能力。在某些情况下，可能会要求后勤人员去采购，并使用专用的测试产品以进行运行前或运行后的检查。DET（Probability of Fault Detection，故障检测概率）的计算公式为

$$\mathrm{DET}=\frac{\text{检测到真实故障}}{\text{真实故障总数}} \tag{14.7}$$

4. 故障隔离概率

如图14.1所示，对于给定百分比的维修行为，故障隔离概率（Probability of Fault I-solation，ISO）与如何设计系统、子系统、设备或组件，以便隔离出一个或多个低级别要素的故障相关。对于给定组件，

$$\mathrm{ISO}_n=\frac{\text{隔离出 } n \text{ 个部件的维修行为数}}{\text{维修行为总数}} \tag{14.8}$$

对于某给定产品，由于产品处于一个不会失效的模糊组内，所以它会经历一定数量的拆除行为或维修行为。ISO也可以定义为此拆除或维修行为的数量除以维修行为总数：

$$\mathrm{ISOL}_i=\frac{\begin{array}{c}\text{由于包含在模糊组内而引起的对产品 } i\\ \text{的(无故障)拆除行为数或维修行为数}\end{array}}{\text{产品 } i \text{ 的维修行为总数}} \tag{14.9}$$

就可测试性的设计而言，ISO和ISOL规定了模糊组的大小。从后勤人员和维修人员的观点出发，维修人员必须要拆除一个或更多个子组件，这将引起更多的维修时间（停机时间）。如果设计允许对一个产品或组件进行足够长时间的故障隔离，后勤人员还必须提供更多产品备件或组件，以允许对产品保障和人力更长时间的利用。

模糊组的拆除和更换也会影响已拆除部件的无法重现率。例如假定在同一时间内只能有一个部件失效，三个印制电路板（PWB）引起一个发送器失效的情况，除非维修人员或车间已经知道运行良好的子组件，否则，用于替换的PWB将必须由更高级别的维修提供，或向制造商订购。维修人员一次只能订购一块PWB，拆除已安装的PWB，并安装上新的，以此来检查是否解决了问题。如果问题没有解决，维修人员就要订购清单上的下一块PWB，或者拆下新的PWB，重新装上旧的。如果这次维修成功，按照通常的维修计划，将对失效产品进行维修或是报废。也许维修人员会同时购买全部三个PWB，并通过替换来进行故障隔离。在这种情况下，维修人员要将两个运行良好和一个发生故障的PWB返回给产品供应商或制造商。因此，运行良好（无故障）产品作为模糊组中的一部分，导致的结果是它的拆除可能会在下一个维修级别变成CND。

5. 维修率

维修率（Maintenance Action Rate，MAR）是指在单位时间内，每个运行单元维修行为的数量。一般来说，MAR由式（14.10）计算：

$$\begin{aligned}\mathrm{MAR}=&\text{单位时间内检测到的真实故障数}+\\&\text{单位时间内无法执行的检测行为数}+\\&\text{单位时间内由模糊组引起的拆除行为数}\end{aligned}$$

$$= DET \times \lambda_L + CND \times MAR + ISOL \times MAR$$
$$= DET \times \lambda_L / (1 - CND - ISOL) \tag{14.10}$$

其中，CND 是维修行为变成一个不可重现事件的概率，ISOL 是产品作为模糊组的一部分所引起的维修行为概率。在式（14.10）中，λ_L 是级数或后勤故障率。此处没有考虑虚警，因为它定义的是维修而不是故障或故障指示。

某一维护级别上的 ISOL 行为（也就是移除子组件的级别）会在下一维修级别转化为 CND 行为，则

$$MAR = DET \times \lambda_L / (1 - CND) \tag{14.11}$$

如果所有的拆除或维修行为处理的都是真实故障，或者在有物资供应的情况下，没有对产品进行更深入的检查来决定它是否存在故障，就将它拆除或送至更高的级别进行维修，那么，CND 等于 1，且 MAR 由式（14.12）计算

$$MAR = DET \times \lambda_L \tag{14.12}$$

如式（14.12）所定义，CND 和 ISOL 并不是统计独立的，且不能同时为 1。

在某些时候，冗余产品中的失效如果不是立即修复的话，计算后勤资源需求时一般不考虑冗余。对于大多数带有冗余产品的系统而言，当冗余产品中的某一个失效时，要么立即开始维修，要么在不妨碍运行时维修。然而对于重型冗余产品来说（n 中取 m 冗余系统必须有效），可以延迟维修，直到一定数量的冗余产品失效，或计划（周期性）维修事件发生。在后一种情况中，产品后勤保障仍然要（同时）提供与即时维修同样的零件数量、技术数据、产品支持以及同样的维修人员。尽管如此，在 n 中取 m 冗余系统延迟维修的情况下，后勤人员可能要在同样的时限内，将产品恢复到功能完备状态，就像只有一个冗余子系统失效一样。如果要在失效发生时进行维修，此时限要求后勤人员计划并提供比预先要求更多的维修人员及产品支持。

6. 需量率

对于某给定时间周期（Time Period）T_L，需量率（Demand Rate，DEM）是能转换为绝对预期维修行为数量的维修率。根据式（14.2），在时间周期 T_L 中的维修行为预期数量，即需量率（DEM）由式（14.13）计算：

$$DEM = MAR \times T_L \tag{14.13}$$

时间的单位必须和 MAR 的单位一致。如果 MAR 是依照每个运行小时内的失效、拆除或维修行为定义的，那么时间也必须以运行小时为单位；如果 MAR 以行程小时为单位，那么时间也必须采用同样的单位。例如为了确定运行（组织）现场内消耗性零件的预期需求，后勤人员需要将 T_L 定义为

$$T_L = K_u \times N_{sys} \times OPHRS \times RESUP \tag{14.14}$$

式中 K_u——设备利用率，某个运行时间单位（例如飞行小时）与其他单位（例如运行小时）之间的换算系数；

N_{sys}——所支持的产品数；

OPHRS——每个日历时间单位（例如每天）内运行或行程小时数；

RESUP——再补给时间（Re-Supply Time），从外场来源订购零件到收货之间所需的时

间周期。

某些产品的运行率可能与包含运行时间记录设备（例如时间指示器）的父级产品的运行率不同，设备利用率（K_u）的计算公式能够说明此类产品的运行率。例如机载航空电子设备除了要在飞行时运行之外，他们还要在飞行前后的检测中运行。或许某些航空电子设备系统只在扩展应用中的短暂时间内运行。设备利用率还可以用来解释维修行为期间的运行时间，它还能将不同应用情况转换成公共时基（Common Time Base）。

7. 平均故障间隔时间

与后勤资源需求量相关的时间周期可以广义表述为供应保障。通常，把这个概念称为平均故障间隔时间（Mean Downtime，MDT）或平均后勤故障间隔时间（MLDT）——此处称其为MDT。平均故障间隔时间（MDT）是对产品后勤保障做出反应的预期时间，例如某产品可在用户现场进行维修，其MDT由式（14.15）计算：

$$MDT = P_{os} \times [P_{sos} \times (MTTR_{os} + MADM_{os}) + (1 - P_{sos}) \times (MTTR_{os} + MADM_{os} + RESUP)] + (1 - P_{os}) \times TAT_{off} \tag{14.15}$$

式中 P_{os}——能够在现场完成必要维修的概率；

P_{sos}——假定可以在现场完成维修，它是现场存在所需备件的概率；

$MTTR_{os}$——现场维修的平均维修时间；

$MADM_{os}$——现场维修的平均故障间隔管理时间，包括从现场仓库取得备件所用的时间；

RESUP——从外场资源处取得具有最长交付时间的零件所需要的平均时间；

TAT_{off}——使用外场修理设施时，产品装运并维修的周转时间。

当发生故障的产品不能在现场维修，MDT的计算公式如下所示：

$$MDT = 0.0 \times [P_{sos} \times (MTTR_{os} + MADM_{os}) + (1 - P_{sos}) + (MTTR_{os} + MADM_{os} + RESUP)] + (1 - 0.0) \times TAT_{off} = TAT_{off} \tag{14.16}$$

案例14.1阐明了在不同维修方案下MDT的计算过程。与将在第14.3.2节中讨论的一样，MDT的计算公式也适用于物资准备和供应保障需求的确定。

案例14.1　平均后勤故障间隔时间

一个名叫布朗的农民拥有一台拖拉机，为了节省资金，他自己修理拖拉机。他有一个相当不错的车间，可以处理拖拉机出现的大多数故障，但由于零件的成本较高，他手边没有太多的备件。因为住在距离拖拉机经销商较远的偏僻地区，所以他通过电话订购了一些必要的零件，并让经销商运送至维修现场。假设在现场能够完成维修的概率P_{os}是0.90；现场具备必要备件的概率P_{os}是0.40；现场维修的平均维修时间$MTTR_{os}$是6.5h；现场维修的平均故障间隔管理时间$MADM_{os}$是0.5h；从拖拉机经销商取得拥有最长交付时间的零件所需要的平均时间RESUP是2.5d；使用外场修理设备将产品装运并维修的周转时间TAT_{off}是0.5个月。如果拖拉机发生故障，那么拖拉机无法工作的平均时间是多少？

把这些数据代入到式（14.15）中，则

$$MDT = 0.9 \times [0.4 \times (6.5h + 0.5h) +$$

$$
\begin{aligned}
&(1-0.4)\times(6.5\text{h}+0.5\text{h}+2.5\text{d}\times 24\text{h/d})]+\\
&(1-0.9)\times(0.5\text{月}\times 30\text{d/月}\times 24\text{h/d})\\
&=74.7\text{h}\\
&=3.1\text{d}
\end{aligned}
\tag{14.17}
$$

因此，不管布朗的拖拉机什么时候发生故障，他都能预估到拖拉机不能工作的平均时间是3.1d。

假设布朗拥有足够的资金，他可以购买所有修理拖拉机所需零件。那么，现场能够完成必要维修的概率等于1.00，则

$$
\begin{aligned}
\text{MDT}&=0.9\times[1.0\times(6.5\text{h}+0.5\text{h})+\\
&\quad(1-1.00)\times(6.5\text{h}+0.5\text{h}+2.5\text{d}\times 24\text{h/d})]+\\
&\quad(1-0.9)\times(0.5\text{月}\times 30\text{d/月}\times 24\text{h/d})\\
&=42.3\text{h}\\
&=1.8\text{d}
\end{aligned}
\tag{14.18}
$$

拖拉机不能工作的时间仅仅不到2d，而不是3d。对于布朗来说，备有所有的零件是不是明智的投资呢？

假设布朗的拖拉机出现了平均时间为75天的故障，且有1.8天的MDT，拖拉机的可用度是多少？

$$
A_0=\frac{\text{MTBM}}{\text{MTBM}+\text{MDT}}=\frac{75}{75+1.8}=0.976 \tag{14.19}
$$

14.3.2 供应保障——维修配件和消耗品的供应

通常，人们认为Harris（1915）首次提出了库存模型（Inventory Model），Raymond（1931）出版了第一本关于此课题的教材。第二次世界大战迫使军事领域引入了保障研究，这些研究的目的是优化备件和消耗性维修配件的采购和库存（黑盒模型），这些都受到成本、重量、体积或综合因素的约束。在20世纪50年代，零售商和制造商开始接受库存理论，并在减少库存和延期交货的同时，将其作为缩减成本和解决超储的方法。

在这一节中，关于库存模型的讨论仅限于经济采购量（Economic Order Quantity，EOQ），它受制于无短缺（No-Shortage）和允许短缺（Shortage-Allowed）两种情况，并通过模拟一个备用冗余产品的方式供应物资。EOQ问题可用一个确定性方法来解决，而备用模型则要用一种随机方法求解。此处的重点是可靠性对供应保障需求所产生的影响。关注库存模型的读者可以参考文献［Sivazlian and Stanfel，1975］、［Hillier and Lieberman，1970］和［Goldman and Slattery，1967］等。

1. 最优再订购量

一个最简单的库存模型考虑以下情况：假设从库存中以一个恒定的需量率（DEM）来提取产品，并定期补充等量的库存（REPLEN）；同时也假设不允许缺货。建立和维持库存相关的成本包括：

SETUP——在时间周期的起始处建立库存的成本；

UNIT——单元的生产成本；

HOLD——每单位量库存的维持成本。

与发出订单相关的成本由式（14.20）计算：

$$\text{COSTORD} = \text{SETUP} - \text{UNIT} \times \text{REPLEN} \tag{14.20}$$

每个时间周期内的维持成本由式（14.21）计算：

$$\begin{aligned}\text{HOLDCOST} &= \text{HOLD} \times \int_0^{\text{REPLEN/DEM}} (\text{REPLEN} - \text{DEM} \times T)\,dT \\ &= \text{HOLD} \times \text{REPLEN}^2/(2 \times \text{DEM})\end{aligned} \tag{14.21}$$

每个时间周期的总成本由式（14.22）计算：

$$\text{COSTPER} = \text{SETUP} + \text{UNIT} \times \text{REPLEN} + \text{HOLD} \times \text{REPLEN}^2/(2 \times \text{DEM}) \tag{14.22}$$

每单位时间的总成本由式（14.23）计算：

$$\begin{aligned}\text{TOTCOST} &= \text{COSTPER}/(\text{REPLEN/DEM}) \\ &= \text{DEM} \times \text{SETUP/REPLEN} + \text{UNIT} \times \text{DEM} + \text{HOLD} \times \text{REPLEN}/2\end{aligned} \tag{14.23}$$

最优的订购量大小 REPLEN^* 是总成本关于订购量大小的一阶导数，并设定微分方程等于0来计算。因此，最优订购量大小是

$$\text{REPLEN}^* = (2 \times \text{DEM} \times \text{SETUP/HOLD})^{1/2} \tag{14.24}$$

消耗预先给定零件量所需的平均时间或预期时间 TIME^* 是

$$\text{TIME}^* = \text{REPLEN}^*/\text{DEM} = [2 \times \text{SETUP}/(\text{HOLD} \times \text{DEM})]^{1/2} \tag{14.25}$$

考虑到允许存在缺货的情况。假设 STOCK 表示在某一时期开始时的现有库存，每个周期的维持成本 HOLDCOST 由式（14.26）计算：

$$\text{HOLDCOST} = \text{HOLD} \times \text{STOCK}^2/(2 \times \text{DEM}) \tag{14.26}$$

对于给定的缺货惩罚费 SHOR \$，每个时期的缺货成本由式（14.27）计算：

$$\text{SHORT} = \text{SHOR}\ \$ \times (\text{REPLEN} - \text{STOCK})^2/(2 \times \text{DEM}) \tag{14.27}$$

每个周期的总成本由式（14.28）计算：

$$\begin{aligned}\text{COSTPER} &= \text{SETUP} + \text{UNIT} \times \text{REPLEN} + \text{HOLD} \times \text{STOCK}^2/(2 \times \text{DEM}) + \\ &\quad \text{SHOR}\ \$ \times (\text{REPLEN} - \text{STOCK})^2/(2 \times \text{DEM})\end{aligned} \tag{14.28}$$

每单位时间的总成本由式（14.29）计算：

$$\begin{aligned}\text{TOTCOST} &= \text{COSTPER}/(\text{REPLEN/DEM}) \\ &= \text{DEM} \times \text{SETUP/REPLEN} + \text{UNIT} \times \text{DEM} + \\ &\quad \text{HOLD} \times \text{STOCK}^2/(2 \times \text{DEM}) + \\ &\quad \text{SHOR}\ \$ \times (\text{REPLEN} - \text{STOCK})^2/(2\text{REPLEN})\end{aligned} \tag{14.29}$$

为了得到最优再订购量大小 REPLEN^* 以及最优初始库存 STOCK^*，必须得到订购量大小和初始库存大小的一阶偏导数 TOTCOST。令这些导数等于0，对这两个方程同时求解，可推导出以下形式：

$$\begin{aligned}\text{REPLEN}^* &= [2 \times \text{DEM} \times \text{SETUP} \times (\text{SHOR}\ \$ + \text{HOLD})/(\text{HOLD} \times \text{SHOR}\ \$)]^{1/2} \\ \text{STOCK}^* &= \{2 \times \text{DEM} \times \text{SETUP} \times \text{SHOR}\ \$/[\text{HOLD} \times (\text{SHOR}\ \$ + \text{HOLD})]\}^{1/2}\end{aligned} \tag{14.30}$$

消耗再订购零件的平均或预期时间由式（14.31）计算：

$$\text{TIME}^* = \text{REPLEN}^*/\text{DEM} = [2 \times \text{SETUP} \times (\text{SHOR \$} + \text{HOLD})/(\text{DEM} \times \text{HOLD} \times \text{SHOR \$})]^{1/2} \quad (14.31)$$

可能发生缺货的平均或预期时间由式（14.32）计算：

$$\text{SHORTIME}^* = \text{STOCK}^*/\text{REPLEN}^* = \text{SHOR \$}/(\text{SHOR \$} + \text{HOLD}) \quad (14.32)$$

案例 14.2 说明了 EOQ 公式在不允许缺货和允许缺货两种情况下的应用过程。

案例 14.2 经济采购量

你是 Good-Gas 石油公司的库存管理员和采购员。公司的管理高层对于送货的延误和安装在送货车上调节阀的相关成本感到不满。你负责去确定库存中阀门的最优数量，还要决定订购多少阀门才能使公司的成本降到最低。

参照历史数据，可做出以下决定：

SETUP = 100 美元；

HOLD = 5 美元/单位量；

UNIT = 500 美元/单位量；

DEM = 5 单元/月。

将以上数据代入到式（14.30）中，得

$$\begin{aligned}\text{REPLEN}^* &= (2 \times \text{DEM} \times \text{SETUP}/\text{HOLD})^{1/2} \\ &= [2 \times (5\ \text{单元/月}) \times (100\ \text{美元})/(5\ \text{美元/单位量})]^{1/2} \\ &= 14.1\end{aligned} \quad (14.33)$$

消耗预先给定的零件量所需的平均或预期时间可由式（14.25）得出：

$$\begin{aligned}\text{TIME}^* &= [2 \times \text{SETUP}/(\text{HOLD} \times \text{DEM})]^{1/2} \\ &= \{2 \times (100\ \text{美元})/[(5\ \text{美元/单位量}) \times (5\ \text{单元/月})]\}^{1/2} \\ &= 2.8\ \text{月}\end{aligned} \quad (14.34)$$

公司管理高层建议，在减低库存成本的前提下，可以延误少量的送货工作。每次 10 美元的缺货成本是可以接受的。由式（14.30）可得

$$\begin{aligned}\text{REPLEN}^* &= [2 \times \text{DEM} \times \text{SETUP} \times (\text{SHOR \$} + \text{HOLD})/(\text{HOLD} \times \text{SHOR \$})]^{1/2} \\ &= \{2 \times (5\ \text{单元/月}) \times (100\ \text{美元}) \times [(10\ \text{美元/缺货}) + (5\ \text{美元/单位量})]/[(5\ \text{美元/单位量}) \times (10\ \text{美元/缺货})]\}^{1/2} \\ &= 17.3\ \text{单元}\end{aligned}$$

$$\begin{aligned}\text{STOCK}^* &= \{2 \times \text{DEM} \times \text{SETUP} \times \text{SHOR \$}/[\text{HOLD} \times (\text{HOLD} + \text{SHOR \$})]\}^{1/2} \\ &= \{2 \times (5\ \text{单元/月}) \times (100\ \text{美元}) \times (10\ \text{美元/缺货})/[(5\ \text{美元/单位量}) \times [(10\ \text{美元/缺货}) + (5\ \text{美元/单位量})]]\}^{1/2} \\ &= 11.5\ \text{单元}\end{aligned} \quad (14.35)$$

由式（14.31），可计算重新订购零件的平均或预期可使用时间：

$$\text{TIME}^* = \text{REPLEN}^*/\text{DEM} = (17.3\ \text{单元})/(5\ \text{单元/月}) = 3.5\ \text{个月} \quad (14.36)$$

式（14.32）给出了仓储需求出现的平均或预期时间：

$$\text{SHORTIME}^* = \text{STOCK}^*/\text{REPLEN}^* = (11.5\ \text{单元})/(17.3\ \text{单元}) = 0.7 \quad (14.37)$$

2. 备件的可用性及供应

某些企业可以接受现有库存的不足。缺货单位成本 SHOR $ 可用来表示利润损失或与利润损失、客户满意度损失相关的费用因素。在某些运行环境下，很难评估备件不足带来的成本。此时，供应模型常用的方法是认为此问题发生在有备份冗余的产品中。如果假定时间相关的故障率保持不变，那么就可以预先推断能够使用累积 Poisson 方程来为备用冗余产品建模，即

$$P(X \leqslant x) = \sum_{i=0}^{x} \frac{(\lambda \times \text{TIME})^{i} \times \exp(-\lambda \times \text{TIME})}{i!} \tag{14.38}$$

为了以供应模型的形式使用，式（14.38）可改写为

$$P_{s}(X \leqslant S_{in}) \leqslant \sum_{i=0}^{S_{in}} \frac{(\text{DEM}_{in})^{i} \times \exp(-\text{DEM}_{in})}{i!} \tag{14.39}$$

式中 $P_s(X \leqslant S_{in})$——需量（X）等于或小于库存备件的概率，也就是说，在需求产生时，备件可以交付使用的概率；S_{in}是库存中的备件数量。

DEM_{in}——给定时间周期内的预期需求。

如式（14.13）所示，预期需求可以简化为产品故障率（λ）和应用数量（N_1）以及每个应用的平均运行时间（OPTIME）的乘积，即

$$\text{DEM}_{in} = N_1 \times \lambda \times \text{OPTIME} \tag{14.40}$$

式（14.40）也可以记作 MAR（维修率）的一个函数。MAR 可能包括除了真实故障之外的 CND 或维修事件，则

$$\text{DEM}_{in} = N_1 \times \text{MAR} \times \text{OPTIME} \tag{14.41}$$

预期需求也可以用产品故障率或维修率与 MDT 的函数来表示为

$$\text{DEM}_{in} = \text{MAR} \times N_1 \times \text{UTIL}^{\ominus} \times \text{MDT} \tag{14.42}$$

对于供应保障，需要计算备件的 MDT，而保障系统或子系统则不需要。备件的 MDT 可计算为

$$\text{MDT}_i = P_{os} \times [P_{sos} \times (\text{MTTR}_{os} + \text{MADM}_{os}) + (1 - P_{sos}) \times (\text{MTTR}_{os} + \text{MADM}_{os} + \text{RESUP})] + (1 - P_{os}) \times (\text{TAT}_{off}) \tag{14.43}$$

式中 MTTR_{os}——在现场维修失效组件、子组件或模块所需要的平均时间；

RESUP——维修第 i 个产品，现场无可用零件时获得零件所需的时间；

TAT_{off}——不能在现场完成故障产品维修的情况下，获得替换产品所需的时间。

案例 14.3 阐述了式（14.39）、式（14.42）以及式（14.43）作为供应模型的使用情况。

案例 14.3 基本供应问题

你是 Lightning-Overnite 快递公司的职工，你的一项工作职责是为一个拥有 150 辆用来配送小型包裹的载货汽车的新车队供应物资。因为收货日程并不固定，每一辆载货汽车上都配备有双频道无线电通信设备，以通知驾驶员在不同的地点停车接货。

⊖ UTIL 为平均利用率。

如果无线电通信设备不能运行，受影响的载货汽车将不能工作。管理层想要了解：应该储备多少台无线电通信设备，才能保证 95% 的载货汽车离开公司时都配有可用的通信设备。

根据无线电设备制造商提供的数据以及公司文件中以往的维修数据，可以获得以下信息：

MAR = 0.0002 拆除/运行小时；

$P_{OS} = 0.10$；

$P_{SOS} = 0.50$；

$MTTR_{OS} = 2.5h$；

$MADM_{OS} = 0.5h$；

RESUP = 3d；

$TAT_{off} = 14d$；

UTIL = 10h/d。

由无线电设备失效引起的 MDT 为

$$\begin{aligned} MDT = 0.10 \times [0.50 \times (2.5h + 0.5h) + \\ (1 - 0.50) \times (2.5h + 0.5h + 3d \times 24h/d)] + \\ (1 - 0.10) \times (14d \times 24h/d) = 306.0h \end{aligned} \tag{14.44}$$

MDT 中的预期需求计算如下：

$$\begin{aligned} DEM = (0.0002\ \text{拆除/运行小时}) \times (150\ \text{辆载货汽车} \times 1\ \text{无线电设备/载货} \\ \text{汽车}) \times (10h/d \times 1d/24h) \times (306.0h) = 3.8\ \text{个无线电设备故障} \end{aligned} \tag{14.45}$$

所需备用无线电设备数量由式（14.46）计算：

$$P_s(X \leqslant S_e) = 0.95 \leqslant \sum_{i=0}^{S} \frac{(3.8)^i \times \exp(-3.8)}{i!} \tag{14.46}$$

表 14.1 中给出了式（14.46）迭代计算的结果。分析结果表明，在外场修理车间维修，或拆除替换失效无线电通信设备并收到替换品所花费的时间内，可以预计平均故障少于 4 个。然而为了达到管理人员的要求，公司必须要储备 7 台设备。

表 14.1　拥有必要备件的概率

S	P_s
0	0.02
1	0.10
2	0.26
3	0.47
4	0.66
5	0.81
6	0.91
7	0.96

3. 包含可更换零件的产品的供应

案例14.3是一种典型的自顶向下的供应方法。分析人员假设下一组件级别（也就是说，维修零部件可能是子组件、模块、产品或是以上各项的组合）的库存水平，以确定需要备用产品的数量。为了供应子组件、模块以及产品，分析人员假设备件可用且拥有供应响应时间。必须供应下一组件级别的产品，使拥有可用维修零部件的概率总和等于或大于上一组件级别计算模型中所使用的值。

在由底向上的方法中，后勤分析人员要确定最低组件级别上所要保持的库存水平，并以此向上一级递进。在每一个潜在的存储点处，对于MDT和存储数量的计算是基于备件充裕的概率完成的，备件充裕概率在前一保障级别（更低的级别）中决定。

备件的充裕度、可用性或拥有所要求备件的概率是

$$P_{s,sys} = \prod_{k=1}^{m} P_{s,k} \tag{14.47}$$

其中，

$$P_{s,k} = \sum_{j=0}^{S_k} \frac{\mathrm{DEM}_k^j \times \exp(-\mathrm{DEM}_k)}{j!} \tag{14.48}$$

式（14.48）是单元S_k的备件充裕度，DEM_k是在组件的下一级别上，对第k个单元的预期需求。

在确定MDT的过程中，所使用的备件或维修配件拥有概率P_{sos}与$P_{s,k}$是不同的。P_{sos}是假定故障已发生时能从现场库存提取备件的概率。式（14.49）可用来计算P_{sos}的值：

$$P_{sos} = P_{sc} \times P_s \tag{14.49}$$

其中，P_{sc}是所需零件存在于现场库存的概率。假定所需零件存在于现场库存清单中，P_s是当时就有零件的概率。计算P_{sc}和P_s的公式如下：

$$P_{sc} = \frac{\sum_{i=1}^{k} N_i \times \lambda_i}{\lambda_T} \tag{14.50}$$

$$P_s = \frac{\sum_{i=1}^{k} N_i \times \lambda_i \times P_{s,i}}{\sum_{i=1}^{k} N_i \times \lambda_i} \tag{14.51}$$

式中 λ_i——存在于现场库存清单的第i个零件的故障率；

N_i——第i个零件应用的数量；

λ_T——考虑到存在和不存在零件的情况下，所供应单元、组件或子组件的总故障率；

$P_{s,i}$——第i个零件充裕的概率。

$$P_{s,i} = \sum_{j=0}^{S_i} \frac{\mathrm{DEM}_i^j \times \exp(-\mathrm{DEM}_i)}{j!} \tag{14.52}$$

式中 $S_i \geqslant 1$。

因此，P_{sos}也可以表示为

$$P_{sos} = \frac{\sum_{i=1}^{k} N_i \times \lambda_i \times P_{s,i}}{\lambda_T} \tag{14.53}$$

案例 14.4 将描述这个公式的使用方法。

案例 14.4　包含可替换部件产品的供应

假设某产品由 8 个组件构成，表 14.2 给出了它们的故障率、预期需求和备件充裕度。由这些已知数据可得 $\lambda_T = 0.0188$，$P_{sos} = 0.9466$。

$$\begin{aligned} P_{sos} = &(1/0.0188) \times (0.0007 \times 0.9977) + \\ &0.0028 \times 0.9970 + 0.001 \times 0.9998 + 0.006 \times 0.9769 + \\ &0.0075 \times 0.9927 = 0.9466 \end{aligned} \tag{14.54}$$

表 14.2　案例 14.4 的数据

组　件	λ_i	D_i	是否输入	S_i	$S \geq 1$ 时的 $P_{s,i}$
1	0.0007	0.07	是	1	0.9977
2	0.0028	0.28	是	2	0.9970
3	0.0005	0.05	否	0	0.0000
4	0.0002	0.02	否	0	0.0000
5	0.0010	0.10	是	2	0.9998
6	0.006	0.6	是	2	0.9769
7	0.0001	0.01	否	0	0.0000
8	0.0075	0.75	是	3	0.9927

很多维修产品和供应产品都是多层级的，在哪一级别上进行维修，是根据产品的可维修性、可测试性设计、维修原则、维修的经济性确定的。产品供应提供在每个维修级别上维修产品所需的备件和维修配件的范围与深度，并以此对维修方案做出响应。对于军工产品来说，仓库或库存控制站是最高级别的维修和库存管理以及备件和维修配件的实际存储点。仓库从内部或民用产品制造商那里购买备件和维修配件，维持实际仓储控制，并将备件和维修配件分配给下一级别。

项目经理或指定的供应保障专家确定需要采购、输入至库存以及分配给下一级别的备件和维修配件的范围与深度，并将其作为现场库存使用。必须为每一级制定供应模型，并用其来计算每个潜在可维修或可更换零件的存储量；同时还必须为可维修和消耗性产品制定多级模型或独立供应模型。典型的可维修和消耗性产品模型可表示为

$$\begin{aligned} \mathrm{MDT}_{1,i} = [&P_{os,1} \times P_{sos,1} \times (\mathrm{MTTR}_{os,1} + \mathrm{MADM}_{os,1}) + \\ &(1 - P_{sos,1}) \times (\mathrm{MTTR}_{os,1} + \mathrm{MADM}_{os,1} + \mathrm{RESUP}_1) + \\ &(1 - P_{os,1}) \times (\mathrm{TAT}_{off,1})]_i \end{aligned} \tag{14.55}$$

现场（中继级）可维修产品的 MDT 为

$$\mathrm{MDT}_{1,i} = \mathrm{RESUP}_{1,i} \tag{14.56}$$

外场可维修产品或消耗品的 MDT：

$$\mathrm{DEM}_{1,i}=D_{1,i}=\mathrm{MAR}_i\times N_i\times \mathrm{UTIL}_i\times \mathrm{MDT}_{1,i} \tag{14.57}$$

现场需求和库存水平为

$$P_{s,1,i}=\sum_{j=0}^{S_{1,i}}\frac{\mathrm{DEM}_{1,i}^{j}\times\exp(-\mathrm{DEM}_{1,i})}{j!} \tag{14.58}$$

在仓库或主要物资库存控制站处，$\mathrm{MDT}_{D,i}$为

$$\begin{aligned}\mathrm{MDT}_{D,i}=[&(1-Z_i)\times(\mathrm{MTTR}_{os,D}+\mathrm{MADM}_{os,D}]+\\&+(1-\mathrm{P}_{sos,D})\times(\mathrm{MTTD}_{os,D}+\mathrm{MADM}_{os,D}+\mathrm{REORDER}_{com})+\\&+(Z_i)\times(\mathrm{REORDER}_{rep})]_i\end{aligned} \tag{14.59}$$

外场可维修产品或消耗品的 MDT 为

$$\mathrm{MDT}_{D,i}=\mathrm{REORDER}_{com,i} \tag{14.60}$$

其中，下标 D 表示基地级维修或制造商级维修，视具体情况而定。Z_i 是可维修产品的报废率，即可维修产品返回基地后不能进行维修的概率；$\mathrm{REORDER}_{com}$是从制造商那里预定并收到备件的平均时间；$\mathrm{REORDER}_{rep}$是从制造商那里订购并收到新的可修理产品所需要的平均时间。

现场需求和库存水平可由式（14.61）得出：

$$\mathrm{DEM}_{D,i}=D_{D,i}=M_{sites}\times\mathrm{MAR}_i\times N_i\times\mathrm{UTIL}_i\times\mathrm{MLDT}_{D,i}$$

$$P_{s,D,i}=\sum_{j=0}^{S_{D,i}}\frac{\mathrm{DEM}_{D,i}^{j}\times\exp(-\mathrm{DEM}_{D,i})}{j!} \tag{14.61}$$

其中，Msites 指仓库所支持的运行场地数，案例 14.5 和 14.6 说明了以上公式的使用方法。

4. 备件优化

使用动态规划方法，可以很容易地把供应清单优化为一个单独的变量。该方法的使用过程如下：

1）对于每一个维修级别和每个组件、子组件以及产品，确定受以下因素影响的预期需量率（$\mathrm{DEM}_{R,i}$）：

① 维修级别。

a. 现场备件。

b. 仓库或物资库存控制站。

② 组件级别。

a. 产品或线路可替代单元。

b. 组件或车间可替换装置。

c. 子组件或模块。

d. 产品及零件。

2）确定每个产品的单位成本（C_i）。

3）对于特定的维修级别和组件级别（例如产品的现场备件），计算：

$$P_{s,sys}=\prod_{i=1}^{m}\sum_{j=0}^{S_i}\frac{\mathrm{DEM}_i^{j}\times\exp(-\mathrm{DEM}_i)}{j!} \tag{14.62}$$

式中 DEM_i——在下一组件级别上第 i 个单元的预期需量率。

4）为每个产品 i，通过将库存加1来确定缺货风险的变化：

$$DEL-S_i=\frac{DEM_{j+1}\times \exp(-DEM_i)}{j!} \tag{14.63}$$

5）为每个产品 i，确定每消耗1美元缺货风险的变化：

$$DEL-S\$_i=DEL-S_i/C_i \tag{14.64}$$

6）选取拥有最高 $DEL-S\$_i$ 值的产品，为其库存加1。

7）继续执行此过程，直至达到某个资金界限，或 $P_{S,sys}$ 已经达到所要求的级别。

产品可靠性或可靠性中的变量是如何影响供应的？案例14.5是合理改进产品可靠性的示例，此案例中的现场可靠性低于预期可靠性。

案例14.5 供应模型的使用

作为一个修理与维护（R&M）工程师，你需要对一个工程修改建议做出评估。此建议的内容是对一个“强化的”无线电通信设备中使用的产品进行升级。任务的一部分是评估为改进可靠性而要对备用无线电通信设备做出的修改。

当前产品的平均失效间隔时间（MTBF）是1500运行小时。假定改进后的MTBF为1950运行小时，如果修改设计建议被采纳，那么可靠性会有30%的提升，则

$$MAR=W\times\lambda_{sys}/(1-CND) \tag{14.65}$$

式中 MAR——维修率；

$W=1.00$；

$\lambda_{sys,old}=1/MTBF=$（1/1500）故障/运行小时 $=0.0006667$ 故障/运行小时；

$\lambda_{sys,new}=1/MTBF=$（1/1950）故障/运行小时 $=0.0005128$ 故障/运行小时；

$CND=0$。

因此，

$MAR_{old}=0.0006667$ 故障/运行小时；

$MAR_{new}=0.0005128$ 故障/运行小时。

令

$$DEM=MAR\times N_{sys}\times OPHRS\times MLDT \tag{14.66}$$

且

$$\begin{aligned}MLDT=&P_{os}\times[P_{sos}\times(MTTR_{os}+MADM_{os})+\\&(1-P_{sos})\times(MTTR_{os}+MADM_{os}+RESUP)]+(1-P_{os})\times(TAT_{off})\end{aligned} \tag{14.67}$$

令

$N_{sys}=30$ 个产品；

$OPHRS=10$ 运行小时/（产品·天）；

$P_{os}=0.90$；

$P_{sos}=0.90$；

$MTTR_{os}=6.5h$；

$MADM_{os}=0.5h$；

RESUP = 2.5d；

TAT_{off} = 0.5 个月。

则

$$DEM_{old} = (0.0006667\text{ 故障/运行小时}) \times (30\text{ 系统}) \times (10\text{ 运行小时/(产品·天)}) \times (1d/24h) \times \{0.90 \times 0.90 \times (6.5h + 0.5h) + (1-0.9) \times [6.5h + 0.5h + 2.5d \times (24h/d)] + (1-0.9) \times [0.5\text{ 月} \times (30\text{ 天/月}) \times (24h/d)]\} + (0.0006667\text{ 故障/运行小时}) \times (596.25\text{ 运行小时}) = 0.396\text{ 个故障} \quad (14.68)$$

$$DEM_{new} = (0.0005128\text{ 故障/运行小时}) \times (596.25\text{ 运行小时}) = 0.306\text{ 个故障} \quad (14.69)$$

使用下面的 Poisson 方程可以计算得出备件水平。假定所要求的 P_s 为 0.95，则

$$P_{s,old} = 0.95 \leqslant \sum_{j=0}^{S_{old}} \frac{(DEM_{old})^j \times \exp(-DEM_{old})}{j!} = 0.95 \leqslant 0.673 + 0.266 + 0.053 = 0.992 \quad (14.70)$$

$$S_{old} = 2$$

$$P_{s,new} = 0.95 \leqslant \sum_{j=0}^{S_{new}} \frac{(DEMAND_{new})^j \times \exp(-DEMAND_{new})}{j!} = 0.95 \leqslant 0.736 + 0.225 = 0.961 \quad (14.71)$$

且 $P_{s,new} = 1$。

所采纳的设计变更将节省 1 个备件的成本。当 P_s 为 0.90 时，能节省多少成本？当 P_s 等于 0.90 时，旧产品和新产品（改进后的设计）都只需要 1 个备件。

案例 14.6　可替换组件的备件

假定某无法维修模块的 MTBF 是 4000 运行小时/故障，有 15 个运行站点和 1 个仓库。每个运行站点为 50 架飞机提供保障。每架飞机使用两个模块，每个月飞行 100h。中继级的重新补给时间是 2 周。从制造商处获得模块的交付时间是 15 个月。如果要求备件充裕度为 0.90，那么每个运行站点及仓库的备件水平如何？

在每个运行站点的预期需量率为

$$DEM_{opsite} = (1/MTBF) \times NACFT \times NUNITS \times OPHR \times SETUP = [1/(4000\text{ 运行小时/故障})] \times (50\text{ 架飞机/站点}) \times (2\text{ 单元/飞机}) \times (2\text{ 周} \times 1\text{ 月/4 周}) = 1.25\text{ 个故障/站点} \quad (14.72)$$

为使运行站点的备件充裕度达到 0.90，所需的备件水平可使用累积 Poisson 方程计算：

$$P_s = 0.90 \leqslant \sum_{j=0}^{S_{onsite}} \frac{(DEM_{opsite})^j \times \exp(-DEM_{opsite})}{j!} \leqslant 0.286 + 0.358 + 0.224 + 0.093 = 0.961 \quad (14.73)$$

且 $S_{onsite} = 3$ 备件/站点。

基地级维修的预期需量为

$$DEM_{depot} = NSITES \times (1/MTBF) \times NACFT \times NUNITS \times OPHR \times REORDER$$

$$= (15\text{ 个站点}) \times [1/(4000\text{ 运行小时/故障})] \times (50\text{ 架飞机/站点}) \times (2\text{ 单元/飞机}) \times [100\text{ 运行小时/(飞机}\cdot\text{月)}] \times (15\text{ 个月}) = 562.5\text{ 个故障} \quad (14.74)$$

在基地级维修上来说，模块是消耗品或是废品。仓库必须储备足够数量的模块，以便在重新订购周期内满足预期需求。

如果不使用计算机来计算累积 Poisson 方程，那么在计算预期需求的备件水平时，所得结果将和本例中的数值一样冗长。使用式（14.75）可以推导出一个近似数值：

$$S = \text{DEM} + Z_\alpha \times (\text{DEM})^{1/2} + Z_\alpha^2/8 \quad (14.75)$$

其中，Z_α 是所需备件充裕度的正态变量。例如本例中的 $P_s = 0.90$，$Z_\alpha = 1.29$，则 $S = 594$。

什么是经济采购量（EOQ）？假设仓库的建立成本是2500美元，库存维持费用是每备件25美元，使用式（14.2），经济采购量为

$$\text{REPLEN}^* = (2 \times \text{DEM} \times \text{SETUP/HOLD})^{1/2} = (2 \times 594 \times 2500/25)^{1/2} = 345 \quad (14.76)$$

消耗现有库存零件的预期时间（即订单间隔时间）可由式（14.11）计算得出：

$$\text{TIME} = \text{REPLEN}^*/\text{DEM} = 345/594 = 0.58 \quad (14.77)$$

按照日历时间计算：

$$\text{TIME}^* = 0.58 \times 15\text{ 个月} = 8.7\text{ 个月} \quad (14.78)$$

案例的结果如下：

① 必须要采办拥有594个单元的初始备件库。

② 马上订购345个单元，并保证在飞机开始运行后15个月内交货。

如果现场需求和预期需求一致，在345个单元投入使用之后（大约8.7个月），订购另外345个单元，必须在15个月内（第23.7个月）送达；如果现场需求与预期需求有较大的差异，那么必须重新计算备件水平和经济采购量，且购置数量也要做出相应的调整。

受实际需求率统计中显著变化的影响，不管345个单元什么时候投入使用，距最后一个订单8.7个月时，重复执行此过程。

14.3.3 人力与人事计划——人员编制

表14.3介绍了典型的多层级修复性维修的人力任务。表中的每个模块表示一个或多个维修、补给、生产制造、质量保证或其他管理型的后勤保障工作。这些必须完成的工作最终使失效产品恢复正常使用，并维修或替换最终产品中的失效部分。每个与后勤保障工作相关的活动都需要专业技术和经验。LSA（后勤保障分析）确认并记录这些基层级、中继级以及基地级维修的人员需求。利用维修性分析得出的数据，LSA还可以确定完成给定维修工作所需的时间（一般为时钟小时，有时也可能是工时）。RLA（维修级别分析）确定将在给定维修级别上完成的维修工作。在ILS（综合后勤保障）规划过程中，可靠性如何影响对人力与人事的需求？人力与人事计划需求是如何估算的？

表14.3 多层级后勤保障产品的任务流程

基层级
在基层级，产生的故障将被隔离至一个线路可替换单元（LRU）。将产生失效的LRU从产品中拆除并用备用LRU进行替换 检测产品是否正常运行 将失效的LRU送至中继级车间进行维修
中继级
在中继级，将LRU隔离至车间替换单元（SRU），以此对其进行维修。拆除失效SRU并用备用的SRU进行替换 监测维修后的LRU是否正常运行 一旦完成LRU的维修，将其返回到基层级，或送至物资库存控制站或存储点 如果没有发现故障，也会将LRU送至基层级，或物资库存控制站或存储点。偶尔也会将中继级不能修理的LRU送至仓库进行维修
基地级
在仓库中，通过将SRU（有时是LRU）故障隔离至部件，以此对其进行维修。拆除并更换失效的部件 检测SRU（或LRU）是否正常运行 一旦完成维修，维修单元将返回至中继级或基地级的物资库存控制站或存储点

使用与确定备件预期需求类似的方法，可以评估人力资源的需求。设计及研发程序的部分工作是确定部件、模块、子组件、组件、单元以及产品的故障率，并将其作为任务发生率并入LSA。补充的人力工时用来完成维修工作，任务发生率为指定技能和经验类型计算每运行小时所需的工时。使用运行剖面（也就是每日历周期中每个单元的运行小时数）和需要保障的单元数量，可以得到特定技术和经验类型的每日历周期（Calendar Period）人力工时。

对专业技能和经验水平或两者兼具的每日历周期内的预期需求工时由式（14.79）计算：

$$\mathrm{MANTIME}_{ijkl}=M_{\mathrm{sites}}\times \mathrm{MAR}_{ij}\times \mathrm{NUNITS}_{i}\times \mathrm{MANTTR}_{ijkl}\times \mathrm{UTIL}_{i} \tag{14.79}$$

式中 $\mathrm{MANTIME}_{ijkl}$——为了保障第i个产品的维修活动j，所需的技能级别k和经验水平l的每日历周期内的工时；

M_{sites}——维修设备保障的运行场地数量；

MAR_{ij}——产品i的维修活动j的维修率（即维修数/单位运行小时）；

NUNITS_{i}——在每个运行站点中第i种单元的数量；

MANTTR_{ijkl}——第i个产品的每次维修活动j所消耗的技能级别k和经验水平l的平均工时；

UTIL_{i}——第i个产品的平均利用率，每个日历周期内每单元的运行小时数；

注意式（14.79）与供应和补给保障公式的相似性。此处，MDT被替换为MANTTR_{ijkl}。

案例 14.7 阐明了这个公式的应用。

案例 14.7 人力需求

拆除及更换小型通勤或商务用喷气机的发动机需要有技术和经验的机械技师、助手以及质量保证检查员。每个拆除操作分别需要 2.5 个工时、1.25 个工时、0.5 个工时。发动机的维修率是每百万飞机单元的飞行小时内 25.0 个拆除操作（也就是每百万飞机运行小时内 50.0 个拆除操作）。现只有 1 个维修场地，需要保障 40 架飞机，每架飞机有 2 个发动机，共计 80 个发动机。每架飞机每个月平均有 110 个飞行小时。在一个月或一年中，与发动机拆除及更换相关的每个技能类别的人力利用率如何？

每个技能级别所消耗的预期工时由下式计算：

$$\begin{aligned}\text{MANTIME}_{ijkl} &= M_{\text{sites}} \times \text{MAR}_{ij} \times \text{NUNITS}_i \times \text{MANTTR}_{ijkl} \times \text{UTIL}_i \\ &= (1\ \text{场地}) \times [25 \times 10^{-6}\text{拆除操作}/(\text{飞机}\cdot\text{单元}\cdot\text{飞行小时})] \times \\ &\quad (40\ \text{架飞机}/\text{站点}) \times (2\ \text{单元}/\text{飞机}) \times [(2.5+1.25+0.5 \times \\ &\quad 110\ \text{飞行小时}/(\text{月}/\text{飞机})] = (0.22\ \text{拆除操作}/\text{月}) \times \\ &\quad (4.25\ \text{工时}/\text{拆除操作})\end{aligned} \tag{14.80}$$

表 14.4 给出了此案例中 MANTIME_{ijkl} 的值。要注意的是：发动机的 MAR 由每飞机单元运行小时内的拆除动作数计算。如果 MAR 是依据发动机运行周期或发动机运行时间计算的，那么就需要用一个换算系数来统一单位（也就是每日历时间内的拆除动作数）。

表 14.4 MANTIME_{ijkl}

日历周期	机械技师	助手	检查员	总计
月	0.55	0.28	0.11	0.94
年	6.60	3.30	1.32	11.22

式（14.79）用来计算某个单元或产品与所执行某种形式维修工作的技能级别相关的工时。通过求指数 i 与 j 的和，可以确定在技能级别 k 和经验水平 l 上的每个日历周期内的总工时。

由此可以得到保障产品必需的预期总工时。这个案例说明维修过程需要三种技能类别，但是只需要一种维修工作类型。

一般来说，备件和所需人员与物资供应一样，只能取整数值。但与物资供应不同的是，兼职人员和全职人员的加班工作可以抵消需求的激增。

在使用由维修性分析或后勤保障分析得出的维修工时数据时，分析人员必须注意，这些数据表示的是进行具体任务时必要的实际工时。通常，维修性分析不包括与维修相关的存取时间、准备工作、故障或者其他周边任务。另外，必须使用人力效率因素将预期实际维修工时转换成雇佣工时，而且维修性或后勤保障分析一般不会给出与维修保障人员（例如管理、质量保证和保障人员等）相关的工时。

为了得出产品的人力与人事需求，从事人力规划工作的后勤分析人员或 R&M 工程师必须要考虑以下因素：

1）进行实际维修或维修保障的可用工时缩减与多项因素相关，包括：

① 闲置时间。

② 直接管理时间。

③ 与周边维修子任务相关的直接维修时间，它在预期或度量任务时间之外。

④ 病假。

⑤ 节日。

⑥ 周末。

⑦ 轮班（每个工作日的可用时间）。

⑧ 工作熟练程度。

⑨ 额外任务或间接任务。

⑩ 工作制度的限制。

2）用更高的百分位数（例如2σ或3σ）值替代预期（平均）值，正态变量的代数应用将为满足人力需求提供更多的保障。可以将这种方法用到 MAR、MANTTR 和 OPTIME 等变量中。

3）可能需要超编人员或加班工作，才能满足均值公式中未涉及的波动或其他瞬时情况。

14.3.4 保障及测试设备——利用率和生产率

用以确定保障及测试设备（Support and Test Equipment，STE）利用率k的均值公式或预期值的方程与确定人员需求的公式类似。STE 与人员一样，只能取整数值，但额外的工作轮班可以满足激增需求。与人员需求不同的是，影响生产率的间接因素或其他因素更加明确，并可以更容易地对其进行预测。对于由产品i的主要组件或更低级别组件的维修行为j引起的 STE，计算其某一项利用率的基本公式如下：

$$\begin{aligned}\mathrm{SETIME}_{ijk} = {} & M_{\mathrm{sites}} \times \mathrm{MAR}_{ij} \times \mathrm{NUNITS}_{i} \times (\mathrm{MTTR}_{ijk} + \mathrm{SET}_{ijk}) \times \mathrm{UTIL}_{i} \times \\ & (1 + \mathrm{MAR}_{k} \times \mathrm{MLDT}_{k}) \times (1 + \mathrm{PMTIME}_{k} \times \mathrm{PMRAT}_{k} + \\ & \mathrm{CALTIME}_{k} \times \mathrm{CALRAT}_{k})\end{aligned} \tag{14.81}$$

式中 SETIME_{ijk}——由主要产品中第i项上，第j个维修活动引起的第k个 SP（备件）的 STE 利用时间；

MTTR_{ijk}——产品i的第j个维修活动使用第k个 STE 的有效维修时间，即 STE 运行小时数；

SET_{ijk}——与保障第i个主要产品的第j个维修活动的第k个 STE 相关的准备时间或其他直接使用时间；

MAR_{k}——第k个 STE 的维修率；

MLDT_{k}——第k个 STE 的每个 STE 维修活动的平均后勤间隔时间；

PMTIME_{k}——对第k个 STE 进行预防性维修（PM）的平均时间；

PMRAT_{k}——第k个 STE 的预防性维修率；

$\mathrm{CALTIME}_{k}$——对第k个 STE 进行校准的平均时间；

CALRAT_{k}——第k个 STE 的校准率。

当使用式（14.81）时，必须要确保时间单位的统一。

对 i 和 j 进行求和，式（14.81）可计算出第 k 个 STE 的总利用率，它与供应保障及人力的需求公式相似。在式（14.81）中，与供应保障公式中定义 MDT 等同的术语定义了每个主要产品、保障产品或校准维修活动所使用保障产品的时间长度。

式（14.81）假定产品保障的预防性维修和校准是产品运行的函数，而不取决于日历间隔时间。对于很多 STE 来说，情况并不是这样，PM 和校准在固定的日历间隔内进行，不考虑它们在日历间隔时间的使用。对于这种情况，式（14.81）可以改为以下形式：

$$\begin{aligned} \mathrm{SETIME}_k = {} & M_{sites} \times \sum_{i=1}^{N_{SE}} \sum_{j=1}^{M_i} \mathrm{MAR}_{ij} \times \mathrm{NUNITS}_i \times (\mathrm{MTTR}_{ijk} + \mathrm{SET}_{ijk}) \times \mathrm{UTIL}_i \times \\ & (1 + \mathrm{MAR}_k \times \mathrm{MLDT}_k) \times \mathrm{CALEND} + \\ & \mathrm{PMTIME}_k \times \mathrm{NPER}_k + \mathrm{CALTIME}_k \times \mathrm{NCAL}_k \end{aligned} \tag{14.82}$$

式中 N_{SE}——第 k 个 STE 支持的主要产品数量；

M_i——对第 i 个主要产品的不同维修活动数量，此产品为了测试支持而使用第 k 个 STE；

CALEND——需要考虑的日历规划周期数量；

NPER_k——CALEND 周期中 PM 的循环次数；

NCAL_k——CALEND 周期中校准的循环次数。

式（14.82）为第 k 个 SE 提供了每个日历规划周期内的利用率（小时）。与人力规划类似，当使用由平均值公式计算得出的 SE 利用小时数时，分析人员必须确定在日历周期或规划周期内 SE 的最大有效小时数。例如某公司标准的工作时间是每天 8 小时轮班，每周 5 天，每年 50 周，每个保障设备每年最多可以运行 2000h。如果式（14.82）得出的结果为 4700 个 STE 运行小时，其中包括 STE 的维修和校准的停机时间，那么就需要 2.35 个单元，取整数的话是 3 个单元。如果正常规定是两个轮班，那么所需 STE 的数量将会减少。

与供应和人力规划一样，为了得到正确的结果，计算中必须保证单位的统一。尤其是当类似 MLDT 的变量出现在公式中时，必须要保证单位的统一，因为产品停机时间、日历时间以及产品可用时间或运行时间必须在共同的时间尺度下度量。而且，前面给出的平均值公式没有考虑到激增需求，也没有考虑满足主要产品的周转时间需求或工作可用性需求。

14.4 维修等级分析

维修等级分析（Repair Level Analysis，RLA），有时也称为维修级别分析，它是一种经济性分析，用来确定产品是需要维修还是报废；除此之外，它还用来确定进行维修或报废工作的维修级别（例如基层级、中继级或基地级）。RLA 是一种迭代分析，它与设计过程互相影响。对于初始设计方法，RLA 用来确定是否可以对某组件进行有效的维修。一方面，如果认定一个产品需要报废，通过去除测试点来简化设计，这样可以节省资金；另一方面，如果初始分析表明一个产品是可修理的，那么重新设计，以增加更

多测试点或初始化环路是合理的方法。表14.5给出了RLA在产品生命周期中的应用情况。

表14.5 RLA与产品生命周期

产品的生命周期阶段	RLA的功能	RLA数据和资源
项目启动与概念探索	开展以下内容的比较研究： ①维修概念：估计可能存在的保障情况 ②产品保障：新产品或已有产品的保障 对以下内容进行运行效能分析： ①为预算编制进行LCC（全寿命周期管理）分析 ②确定保障性和维修级别的非经济性约束	①从运行现场已有的产品获取R&M和LCC数据 ②预测
设计与开发	①为维修性和可测试性修改设计 ②初步量化，确定对产品保障、设施、人员和主要组件供应的需求 ③制定维修和废弃决策 ④评估所提议设计变更对LCC的影响	①R&M预测 ②LSA ③开发预选成本估计
生产现场初始运行	①制定维修级别决策 ②确定供应需求，包括用户/运行现场备件和维修现场的维修零件库存 ③设计更改，评估所提议设计变更对LCC的影响 ④审查并评估后勤保障产品的效能 ⑤更新供应列表 ⑥评估所提议设计变更对LCC的影响	①LSA ②试验结果 ③R&M预测 ④设计更改提议 ⑤现场维修和成本数据

对于很多采办项目来说，RLA是一种严格的经济分析，然而成本与运行可用性/准备状态（或其他有效性措施）之间的权衡分析是很容易执行的。通常，由RLA确定偶生（Nonrecurring）费用和经常性（Recurring）费用，这些成本与所有10个维修和报废后勤元素相关。除非非经济性因素把分析限定为先验性分析（例如印制电路板只能进行基地级维修，密封混合电路则无法维修），否则在每个可以进行维修的级别上，为每个适用的子系统、组件、子组件、模块以及潜在可维修产品计算成本。这些成本的评估要用到后勤保障分析报告（Logistics Support Analysis Report，LSAR）得出的数据以及均值模型。这些模型与前面章节中给出的用于供应、人力以及SE的模型类似。

如果要考虑典型的武器、航空航天产品或电子产品中可维修产品的数量，RLA会变得相当复杂。除非没有预先的经济性的限制，否则就必须要对每个组件、子组件或模块的维修或报废进行评估。对于这两者中的任意一种情况来说，选定的活动可以在三个维修级别中的某一个上完成。

执行 RLA 时，必须要考虑以下这些事项：

① 随机变量［例如 MTBF、维修间隔平均时间（MTBM）、MAR、MTTR 等］的变化会引起维修与报废决策或维修级别的变化。为了评估这些变化的潜在影响，必须要对灵敏度进行研究。

② 与后勤资源相关的成本（例如多用途测试产品或修理设备）必须要分期回报给由特定资源支持的所有产品。如果某个保障产品的维修决策导致这个产品不再需要资源，那么就必须进行迭代分析，并将变化反映出来。

③ 与产品保障相关的成本是非常重要的。如果需要新的或特有的 STE，那么必须要对它们进行评估。必须为此 STE 进行一次 RLA。

④ 尽管组件或子组件的维修是可行的，但对于维修流程合格率和报废率的考虑也是非常重要的。如果维修的合格率较低（高报废率），那么试图进行维修的决策可能会发生变化。

⑤ 运行需求和产品废弃的可能性，或新的可消耗组件的未来不可用性可能会否决以经济为基础的决策（例如当制造商决定不再生产某一组件，但产品中仍然使用这种随时都可得到的组件。如果维修流程、其他技术数据以及必要的 SE 不存在，产品的可用性可能不会提供任何帮助）。

14.5 总结

本章探讨了可靠性，并在有限的范围内探讨了可测试性对备件及维修配件、人员和产品保障的影响。本章首先根据恒定到达率假设提出了保障需求公式，提出了平均故障间隔时间术语；然后提出了 MDT，它反映了产品的维修时间，备件所需的时间，从外场库存运送零件或备件时的产品保障的响应时间，每个维修操作中的人员或产品保障的利用率等；同时也讨论了维修级别分析及其在产品生命周期中的定位。

参考文献

Blanchard, B. S. 1992. Logistics engineering and management. Upper Saddle River, NJ: Prentice Hall.

Goldman, A. S., and T. B. Slattery. 1967. Maintainability: A major element of systems effectiveness. New York: John Wiley & Sons.

Harris, F. W. 1915. Operations and cost. New York: A. W. Shaw.

Hillier, F. S., and G. J. Lieberman. 1970. Introduction to operations research. Ann Arbor: Holden Day, University of Michigan Press.

Raymond, F. E. 1931. Quantity and economy in manufacture. New York: McGraw-Hill.

Sivazlian B. D., and L. E. Stanfel. 1975. Analysis of systems in operations research. New York: Prentice Hall.

第 15 章　产品效能和成本分析

15.1　引言

本章介绍如何将可靠性和维修性数据与产品性能数据结合起来，对产品整体效能进行评估的过程，并介绍如何通过引入成本因素来对最终设计提供更完整的决策依据。首先本章回顾了和产品效能相关的概念；然后为单一模式产品建立了量化效能的广义模型，并将此模型扩展至多模式产品；随后全面讨论了如何分析产品效能；最后讨论了将成本引入决策过程的方法。

我们曾在第 1 章介绍过，产品效能（Product Effectiveness）代表产品满足客户或用户需求的总体能力。它的正式定义为：产品效能是衡量产品在可用性、可信性和能力方面达到预期应用要求程度的指标。

可用性（Availability）——用于衡量产品在开始应用或使用时状态的量，它是产品硬件、相关人员和程序之间关系的函数。

可信性（Dependability）——在运行开始时产品状态已知的情况下，用于衡量应用过程中一个或多个时间点处产品状态的量。

能力（Capability）——在运行过程中产品状态已知的情况下，用来衡量产品满足其应用目标能力的量。能力特指产品的性能范围。产品能力可以用多种形式来衡量，如成功的概率、与最高性能相关的指标、产品的输出（如功率的兆瓦数）或者它的影响（如载货的吨位）。

假如我们在分析一个非常简单的产品，它要么处于“工作”状态，要么处于“非工作”状态，并且在使用过程中无法修复。上述定义将引出下列几个关于效能分析的问题：

① 可用性：在用户需要开始使用产品时，该产品是否处于工作状态？

② 可信性：在整个使用过程中，产品是否一直保持工作状态？

③ 能力：如果产品在整个使用过程中一直处于工作状态，它的表现能否满足功能要求？

为了回答以上问题，我们将“工作”定义为：产品的输出在其设计规格范围内。

令人惊讶的是，这种简单的分析模式完全可以处理很多种情况，至少可以用它进行初步的粗略估计。假设某电视机产品，其“应用”是供用户观看系列足球赛事，产品效能就是用户能够完整看完球赛的概率。那么电视机的效能计算如下：

$$E_{ff} = P\{在赛事开始时,电视机可用\} \times$$
$$P\{电视机能完整播放整个赛事——可用\} \times$$

$$P\{电视机能播出满意的画面和声音——可信\} \tag{15.1}$$

基于这个例子，我们可以扩展出更复杂的情况。例如某个零件超出了规定容限，电视画面出现雪花点，但仍然可以看得见影像，电视机的效能如何？如果电视机没有声音，但是无线电接收装置可用，其效能如何？如果画面色调偏亮，但图像仍然清晰，其效能又如何？这些问题都和产品性能量化方面的“画面和声音效果满意度”相关。如果允许在节目时间段对产品进行维修，我们也可以很容易地提出关于产品可靠性和可信性方面的问题。因此，虽然对产品只进行简单的“工作或非工作”状态的分析具有一定的作用，但我们也需要更全面的分析方法。

15.2 用 Markov 过程量化产品效能的框架

对于一些相对简单、不能维修或只能进行少量维修的产品来说，可用性不是一个很重要的问题，其可信性和能力模型也比较容易建立。但今天的产品已经变得越来越复杂，而且经常会涉及中央计算机、数字传感器、分布式微处理控制器和机内测试能力。这种复杂的可修复产品的实质是产品系统，为其建模的标准技术是采用 Markov 模型。

Markov 过程是由最近历史状态的函数——概率所决定的。Markov 模型是产品状态（例如运行状态和非运行状态）和观测时间的函数。它由一系列概率（p_{ij}）定义，这些概率定义了产品从状态 i 转变到状态 j 的概率。Poisson 过程是一个特殊的 Markov 过程。

为了构建一个 Markov 模型，必须定义产品的所有互斥状态，然后用 Markov 状态方程描述产品从起始状态转变到最终状态的概率。复杂产品的系统模型有很多状态，状态方程求解的计算量也很大，一般都从极其微小的设计点开始着手求解。为了方便计算，一般需要利用合并、近似计算或者 Monte Carlo 模拟等方法来减少状态数量。本小节定义了一个 Markov 模型的框架。关于 Markov 模型近似和简化技术的详细讨论，读者可以参考文献 [Shooman，1990]。

产品状态可以作为描述产品的基础，状态转变可用来反映产品的可靠性和维修性。对于描述产品状态，最简单但不一定最优的方式是首先考虑每个产品部件的有效和失效状态，然后在此基础上描述产品状态。那么，产品状态就是有效部件和无效部件的不同组合。假设一个产品由 n 个部件构成，则该产品共有 2^n 种状态。当部件状态发生变化时（某个有效部件失效，或某个失效部件被维修好），产品状态就会转变。基于某些简化的假设，使用 Markov 过程是相对简单的状态模型分析方法。接下来我们将从可用性、可信性和能力这三个概念出发，对这个模型的几种形式进行描述。

单一运行模式（Single Operating Mode）产品不存在使用中维修（In-Use Repair）。这类产品只有一种工作模式（正如前一小节中对电视机的假设一样），因此，在使用过程中不可能对失效进行修复。在这种情况下，可用以下通用模型来计算产品的效能：

$$E_{ff} = A_v D_p C_{ap} \tag{15.2}$$

式中 A_v——可用性，产品在启动时处于运行状态的概率；

D_p——可信性，产品在整个使用周期内持续运行的概率；

C_{ap}——能力，假定产品在整个使用周期内可信，产品的性能表现。

假设某种应用情况为：在应用时间 t_m 内对产品有持续的性能要求，那么产品效能可以量化为 $E_{ff}(t_k)$ 在时间 t_m 内的平均值，即

$$E_{ff}=\frac{1}{t_m}\int_0^{t_m}E_{ff}(t)\,dt \tag{15.3}$$

需要注意的是：在运行阶段内任何一个采样时间点处，如果状态 j 属于满意状态集合，则产品的能力系数 $c_j=1$；否则，$c_j=0$。上述的 E_{ff} 公式就简化为产品运行状态相对于满意状态的预期比值。

如果 Markov 假设不成立，对于所有的状态转变，产品的能力矩阵写为一个 $N\times N$ 矩阵（N = 产品状态数）。

15.2.1 多功能产品运行的广义模型

假定产品在使用阶段要执行 f 种功能，其中，在区间 t_k 到 $t_k+\delta_k$ 内，该产品执行第 k 种功能，将其记作 T_k。我们将这样的区间称为第 k 个功能区间（Functional Interval）。我们将调用时间间隔 T_{k-1} 到 T_k（也就是从 $T_{k-1}+\delta_{k-1}$ 到 t_k 的时间间隔）之间的第 k 个非功能区间（Nonfunctional Interval）记作 τ_k。图 15.1 中，符号“………”表示功能期（Functional Period），符号“……”表示非功能期（Nonfunctional Period）。

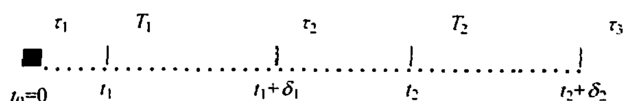


图 15.1 产品在区间 t_k 到 $t_k+\delta_k$ 内执行第 k 种功能

为简单起见，用成功概率来表示产品效能，功能区间没有重叠，且为了表述的完整性，假设功能区间内没有状态转变发生。那么产品效能可表示为

$$E_{ff}=A_v W\pi\,|\,D_p(\delta_k)P(T_k)C_k\,|\,D(\delta_t)P(T_t)C_t \tag{15.4}$$

其中：

$$A_v=[a_1\quad a_2\quad \cdots\quad a_n] \tag{15.5}$$

a_i——在 $t=0$ 时刻，或者产品开始运行时，产品处于状态 i 的概率。

$$W=\begin{bmatrix} W_1 & & \\ & W_2 & 0 \\ 0 & & \\ & & W_n \end{bmatrix} \tag{15.6}$$

W_i——在 $t=0$ 时刻，产品处于状态 i 时，产品被使用的概率。

$$D_p(\tau_k)=\begin{bmatrix} d_{11}(\tau_k) & d_{12}(\tau_k) & \cdots & d_{1n}(\tau_k) \\ d_{21}(\tau_k) & d_{22}(\tau_k) & \cdots & d_{2n}(\tau_k) \\ & \cdots & & \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \tag{15.7}$$

$d_{if}(\tau_k)$—— 在第 k 个非功能区间内，产品由状态 i 变成状态 k 的概率。

$$p(T_k)=\begin{bmatrix} p_1(T_k) & & & \\ & p_2(T_k) & & \\ & & \ddots & \\ 0 & & & p_n(T_k) \end{bmatrix} \tag{15.8}$$

$p_i(T_k)$——如果已知在时件 t_k 处，产品产于状态 i，那么 $p_i(T_k)$ 代表在第 k 个功能区间开始时，时间 $t_k+\delta_k$ 前无状态转变的概率。

$$c_{ap}(T_k)=\begin{bmatrix} c_1(T_k) & & \\ & c_2(T_k) & 0 \\ 0 & & \\ & & c_n(T_k) \end{bmatrix};\ for\ k=1\ to\ f-1 \tag{15.9}$$

$$C_{ap}(T_f)=\begin{bmatrix} c_1(T_f) \\ c_2(T_f) \\ \vdots \\ c_n(T_f) \end{bmatrix} \tag{15.10}$$

$c_i(T_k)$——在第 k 个功能区间，第 i 个产品状态实现所有功能需求的概率。

为了描述的简单性，假设只有两个功能区间，那么

$$E_{eff}=A_vWD_p(\tau_1)P(T_1)C_{ap}(T_1)D_p(T_2)P(T_2)C_{ap}(T_2) \tag{15.11}$$

通常，将该表达式写成 $a_iw_id_{ij}p_jc_jd_{jk}p_kc_k$。它表示产品在状态 $i(a_iw_i)$ 下开始运行，在第一个功能区间开始时转变到状态 j，并且将状态维持至区间（$d_{ij}p_j$）结束；在状态 j（c_j）下成功执行第一种功能；在第一个功能区间后，第二个功能区间（d_{jk}）开始前转变成状态 k；在第二个功能区间（p_k）内维持状态 k；在状态 $k(c_k)$ 下成功地完成第二种功能的概率。

当然，对于一些比较复杂的模型，为了简化分析，可以适当调整一个或多个模型的限制条件。下一小节将对此进行举例说明。

15.2.2 效能评估示例——连续运行

前面所给出模型的一个缺点是它依赖于离散状态点，即 Markov 性。虽然很多产品都可基于此框架进行建模分析，但我们还可以用基本模型概念来分析产品的连续运行。在某些情况下，还可以通过为每个状态转变加上能力度量来完成这种分析。下面的例子详细介绍了这一方法的应用过程。

示例产品的定义：两个通信产品，A 和 B 同时用来传递信息。如果其中任何一个产品失效，另一个产品仍可以单独传递信息（A 和 B 的运行是相互独立的）。在信息传输过程中，任何一个产品的功能失效都是不可修复的，但可以在正常的关闭时间内对产品进行维修。

只要 A 和 B 中任何一个产品可用，信息的传递就会开始。（也就是说，信息的传递不要求 A 和 B 同时可用）。表 15.1 列出了两个产品的平均失效时间间隔，平均修复时间

和信号传输的比特率。为了说明效能的评估方法，我们将使用基本的产品效能模型来回答以下问题：“如果产品效能的定义为：在40min内，信号的传输速率不低于800000比特率概率，那么A和B两个产品组合在一起的效能如何?”在分析过程中，我们假设产品只能有一次状态转变。表15.2用数字表示信号传输过程中的产品状态（字母上方的横线代表失效状态）。

表15.1 产品A和B的平均失效时间间隔、平均维修时间和传输速率

产　品	平均失效间隔时间 /min	平均修复时间/min	传输速率/（bits/min）
A	1200	60	30000
B	2000	80	15000

表15.2 产品状态的表示

配　置	状态编号
AB	1
$A\bar{B}$•	2
$\bar{A}B$	3
$\overline{AB}$	4

可用性计算：产品的可用性（A_v）是指稳态状况下，产品在任何一个时间点能及时运行的概率，可由式（15.12）计算：

$$A_v = \mathrm{MTBF}/(\mathrm{MTBF}+\mathrm{MTTR}) \tag{15.12}$$

特别要指出的是，本例中系统的子产品A和B的可用性分别为

$$\begin{aligned}\mathrm{Avail}(A) &= 1200/(1200+60)=0.9524\\ \mathrm{Avail}(B) &= 2000/(2000+80)=0.9615\end{aligned} \tag{15.13}$$

定义：$a_i = P$（在信号传输开时产品处于状态i——子产品A和B可用性的函数），则

$$\begin{aligned}a_1 &= \mathrm{Avail}(A)\mathrm{Avail}(B) = (0.9524)(0.9615)=0.9158\\ a_2 &= \mathrm{Avail}(A)[1-\mathrm{Avail}(B)] = (0.9524)(0.0385)=0.0366\\ a_3 &= [1-\mathrm{Avail}(A)]\mathrm{Avail}(B) = (0.0476)(0.9615)=0.0458\\ a_4 &= [1-\mathrm{Avail}(A)][1-\mathrm{Avail}(B)] = (0.0476)(0.0385)=0.0018\end{aligned} \tag{15.14}$$

那么，可用性的矢量为

$$\boldsymbol{A}_v = [0.9158 \quad 0.0366 \quad 0.0458 \quad 0.0018] \tag{15.15}$$

可信性计算：因为产品不存在使用中维修，那么可信性仅依赖于子产品A和B的运行可靠性。假设可靠性函数是指数函数，则

$$R(T) = \mathrm{e}^{-t/\theta} \tag{15.16}$$

其中，t是产品运行时间，θ是平均失效间隔时间（Mean Time Between Failure，MTBF）。那么子产品A和B在40min内的可靠性计算如下：

$$R_A(3\text{h}) = \text{e}^{-40/1200} = 0.9672$$
$$R_B(3\text{h}) = \text{e}^{-40/2000} = 0.9802 \quad (15.17)$$

定义：$d_{ij} = P$（从状态变换到状态）$_{3\text{h}}$。表 15.3 列出了状态转变概率。

表 15.3 状态转变概率

配　置	状 态 编 号
$d_{11} = (0.9672)(0.9802) = 0.9481$	$d_{31} = (0)(0.9802) = 0$
$d_{12} = (0.9672)(0.0198) = 0.0192$	$d_{32} = (0)(0.0198) = 0$
$d_{13} = (0.0328)(0.9802) = 0.0321$	$d_{33} = (1)(0.9802) = 0.9802$
$d_{14} = (0.0328)(0.0198) = 0.0007$	$d_{34} = (1)(0.0198) = 0.0198$
$d_{21} = (0.9672)(0) = 0$	$d_{41} = (0)(0) = 0$
$d_{22} = (0.9672)(1) = 0.9672$	$d_{42} = (0)(1) = 0$
$d_{23} = (0.0328)(0) = 0$	$d_{43} = (1)(0) = 0$
$d_{24} = (0.0328)(1) = 0.0328$	$d_{44} = (1)(1) = 1$

能力计算：定义 $c_{ij} = P$（在产品从状态 i 转变到状态 j 的过程中，40min 内至少传输 800000 比特的信号），那么可立刻得出以下结论：对于所有的 $i > 2$，$c_{11} = 1$、$c_{12} = 1$、$c_{22} = 1$、$c_{23} = 0$、$c_{ij} = 0$。

为了描述求解过程，假设 c_{11}、c_{12} 和 c_{22} 都等于 1，因为它们表示在整个产品运行阶段中 A 工作，产品满足传输需求的能力。$c = 0$ 表示产品不能满足传输需求或 A 在整个运行阶段不可用。因为产品 B 不可能在 40min 内传递 800000 比特信号，那么 A 在开始就失效，与之对应的状态下产品的能力为 0。对于 c_{13}、c_{14} 和 c_{24}，情况比较复杂，我们给出以下定义：

① 每分钟的传输速率：r_a 和 r_b 分别表示产品 A 和 B 的每分钟传输速率。

② 产品的运行时间：T。

③ 失效率：λ_a 和 λ_b。

④ 需要传递的总信息量：β_b。

首先考虑 c_{24}，即产品开始时处于状态 2（A 工作，B 不工作），从状态 2 转变到状态 4（A 和 B 都不工作）的过程中，至少传输 β 比特信息的概率。显然，只有 A 在传递 β 比特信息后失效才可能出现这种情况，或者，A 的失效时间 t_a 不早于时间 β/r_a。此概率由式（15.18）计算：

$$c_{24} = \int_{\frac{\beta}{r_a}}^{T} \frac{\lambda_a \text{e}^{-\lambda_a t_a}}{1 - \text{e}^{-\lambda_a T}} \text{d}t_a = \frac{1}{1 - \text{e}^{-\lambda_a T}} \left[\text{e}^{-\frac{\lambda_a}{r_a}\beta} - \text{e}^{-\lambda_a T} \right] \quad (15.18)$$

带入本例中的实际值后，可以计算出 $c_{24} = 0.3296$。

c_{13}的计算方法与此类似。此时，产品状态由 A 和 B 都工作转变到 A 不工作。如果 A 在时间 t_a 处失效，它已传输的信息量为 $r_a r_b$ 比特；因此，仍处于工作状态的 B 将在余下的时间 $T - t_a$ 内传输 $r_b(T - t_b)$ 比特信息。因为这种情况下所要传输的总信息量 $r_a t_b + r_b(T - t_a)$ 至少为 β，所以我们可以定义满足要求的 t_a 下限，并给出相应的概率表达式：

根据提供的数据，我们可推算出 $c_{13}=0.8479$。

能力 c_{14} 表示产品从 A 和 B 都工作到 A 和 B 都失效的状态转变。此处要使用一个卷积方程，它表示在时间段 T 内，A 和 B 传递信号的总和不少于 β 的概率：

$$c_{14}=\int_{\frac{\beta}{r_a}}^{T}\frac{\lambda_a e^{-\lambda_a t_a}}{1-e^{\lambda_a T}}-\mathrm{d}t_a+\int_{0}^{\frac{\beta}{r_a}}\frac{\lambda_a e^{-\lambda_a t_a}}{1-e^{-\lambda_a t}}\int_{\frac{\beta-r_a t_a}{r_b}}^{T}\frac{\lambda_b e^{-\lambda_b t_b}}{1-e^{\lambda_b T}}\mathrm{d}t_b\mathrm{d}t_a \tag{15.19}$$

其中，第一项表示在传递了 β 比特信息后，A 失效的概率；第二项表示在 A 和 B 在时间 T 之前都失效，A 和 B 总共至少传递了 β 比特信息的概率。带入相关的值，我们可以算出 $c_{14}=0.5509$。

通过上述的计算，我们可以得出矩阵 C：

$$C_{(ap)ij}=\begin{bmatrix}1 & 1 & 0.8478 & 0.5509\\ 0 & 1 & 0 & 0.3296\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \tag{15.20}$$

然后就可以利用式（15.21）计算产品效能：

$$E_{ff}=\sum_{i=1}^{4}\sum_{j=1}^{4}a_i d_{ij} c_{ij} \tag{15.21}$$

式（15.21）也可以表示为

$$E_{ff}=A_v\begin{bmatrix}\sum_{j=1}^{4}d_{1j}c_{1j}\\ \sum_{j=1}^{4}d_{2j}c_{2j}\\ \sum_{j=1}^{4}d_{3j}c_{3j}\\ \sum_{j=1}^{4}d_{4j}c_{4j}\end{bmatrix} \tag{15.22}$$

当代入相关值后，我们可以得出：

$$E_{ff}=\begin{bmatrix}0.9158 & 0.0366 & 0.0458 & 0.0018\end{bmatrix}\begin{bmatrix}0.9948\\ 0.9780\\ 0\\ 0\end{bmatrix} \tag{15.23}$$

或 $E_{ff}=0.947$。

15.2.3 模型的适用性

用一个简单的模型去普遍量化诸如产品效能这种复杂的概念是不可能的。如前所述，我们所提出的模型只提供了产品效能分析的概念框架。产品需要通过工作来展示其性能。当它失效时，必须对其进行维修；如果处于运行状态，它们必须实现功能。这是一个模型的本质所在。

所有效能模型的一个共有缺点是它们没有任何能描述复杂产品满足目标要求程度的

单一量。例如在评估一个通信产品时，分析人员可能需要考虑它的电容量、误码率（Error Rate）、安全性以及其他的一些因素。虽然为这些因素提出一个单独的评估量是不可能的，但我们还是可以为每一个重要因素建立一个效能分析量，因此也就产生了一系列的评估量。

实际上，之前所提的模型是有能力满足这一要求的。如果将能力矢量转化为能力矩阵，其中每一列表示与某个特定输出相关的能力（例如电容量、误码率、安全性），那么此模型就有一个矢量解。但只有当所有的可用性和可信性公式都适用于所有的能力量（Capability Measure）时，这种方法才可行。

15.3　产品效能分析所要考虑的因素

前面小节的内容为量化产品效能提供了基础。在本节中，我们将检验在评估设计方案和分析现场产品中所用的方法和所要考虑的因素。图 15.2 是典型效能分析程序的流程图。

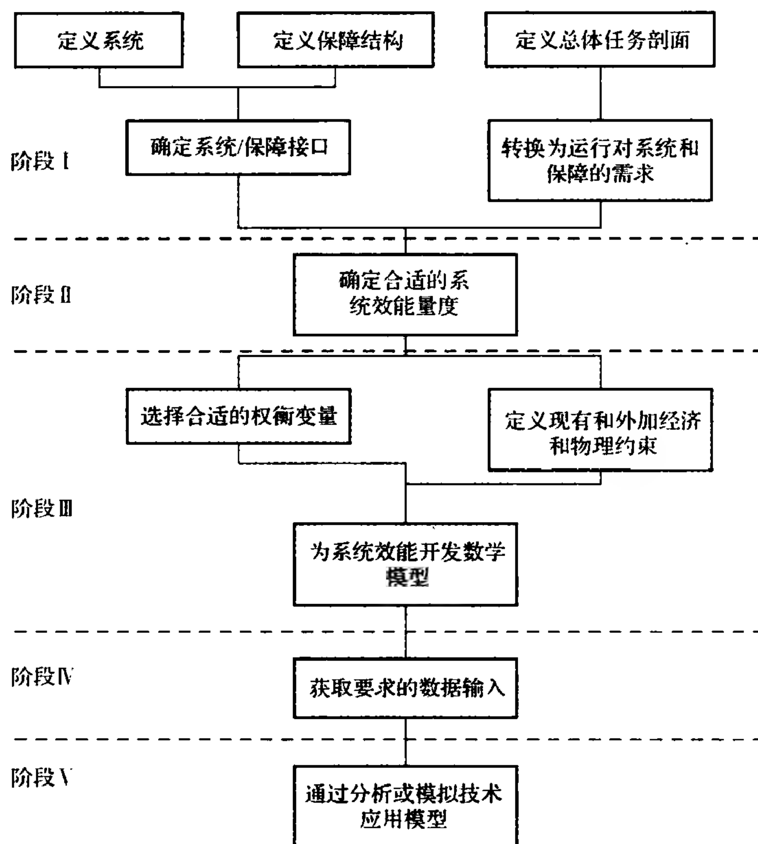


图 15.2　典型的产品效能分析流程

产品效能分析是在多个产品和保障层级上进行的。通常，首先在产品或主要子产品层级（例如计算机系统或数据处理器）和相关的保障层级上进行分析。早期的分析为

总体设计方法提供了决策依据，并为后续的硬件和保障层级分析提供了基础。因此，计算机系统层级的分析有助于定义整体系统框架，并有助于数据处理器层级的分析。此层级的分析将决定中央处理器和相关硬件如何处理输入、计算以及输出。

因为早期阶段的数据通常是有限的，因此必须延迟制定一些决策，另外一些决策则需要根据偶发事件来制定。直到冻结设计之前，产品效能分析每一次迭代的结果都可以用来改进分析模型、分析标准、前期设计以及保障决策等。在产品设计的关键节点，必须要对效能分析进行同步的管理规划。这些步骤和相应节点如图 15.3 所示。分析的过程实际上是把产品需求和约束转化成产品更小部件的参数要求和约束，然后将这些数据提供给相关的设计小组。随着过程的进展，产品部件和产品整体效能的关系将得到更好的定义，这会使得设计方案的决策是建立在客观实际基础之上而非主观基础之上。

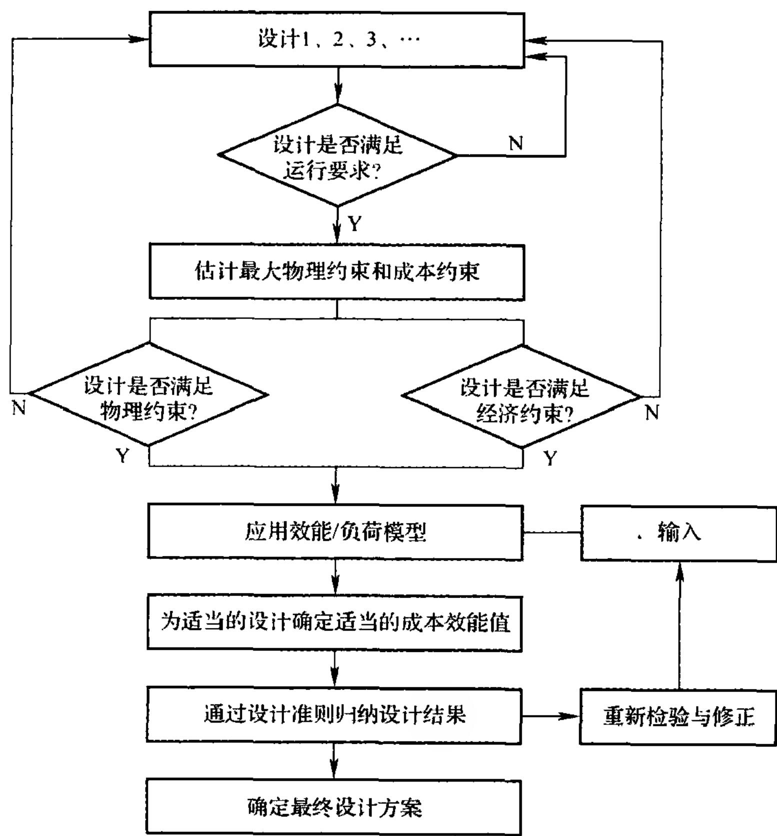


图 15.3 模型实施过程

15.3.1 阶段Ⅰ：定义应用、产品与后勤保障

应用、产品和后勤保障将最终成为产品效能分析模型（阶段Ⅲ）的基础，并且为设计决策标准提供建立依据，阶段Ⅰ将对这三方面进行详细的研究。

此阶段中的第一项任务是将广义的应用目标转化成适合效能分析的定量运行要求。

例如某指挥和控制系统的总体目标——在事故发生后依旧能执行命令——必须转化为具体的要求。这样就会引出一系列的特定要求，例如受到何种类型的攻击后，系统依

然可运行？该产品必须在什么样的环境下运行？可用于系统的和系统所要求的信息数据如何？所有需要的实时数据是什么？

如果系统涉及多种应用，第一阶段将为每种应用产生一系列运行要求。所产生的要求集合应被缩减为一个复合集，来定义产品对主要或最有可能应用类型的有效性。在集合的缩减过程中，可以使用严重度和发生率的权重来为各种应用类型分配权重。

运行要求用于定义产品和主要子产品的边界条件、功能和约束。必须通过初步规范、硬件描述、功能框图等方面清晰定义子产品及其功能，必须对产品和操作者之间的接口、后勤保障功能、所有大型产品以及应用环境进行分析。

在产品开发早期，框图用于分析多模联运能力（Multimodal Capability）、失效影响和冗余等。分析还要包括其他一般的可靠性和维修性设计事项，例如板载测试和模块化航空电子系统的使用。应该在维护级别、可用维修设施和可用维护技能等方面对后勤保障进行检验。为了进一步研究的需要，必须用效能模型（Effectiveness Model）描述获取高可靠性、维修性和准备性的各种方法。

15.3.2 阶段Ⅱ：选择效能量度

阶段Ⅰ中对产品、后勤保障和应用的定义为使用第 1 章和本章前部分所讨论的因素制定效能量度提供了基础。某些因素（如可靠性）适用于所有产品，而其他一些因素只适用于特定产品类型——例如最大起飞总重只和飞机相关。在选择效能量度的过程中，必须注意避免过早地限制设计方案。例如在超声速巡航过程中，一架飞机需要用到低红外线信号，那么就应该根据它来度量效能。如果飞机在缺少再燃装置时进行超声速巡航，此时重新定义量度的话，将会减少留给设计团队的设计选项。

除此之外，还必须定义产品的使用频率和性能度量标准。产品的使用频率很重要，因为产品效能的基础是在特定时间段内满足运行需求或满足其他使用约束（如对于里程数和使用次数的限制）的能力。通常，产品所表现的功能将表明此类时间要求或使用要求的本质，但在某些情况下会有多种选项。表 15.4 以时间为基础，列出了时间要求和使用要求的三种类型。

表 15.4 与产品运行要求相关的时间或使用需求

需求类型	描述	示例
即时需求	在较短持续时间内，产品必须满足给定使用要求	轨道运行卫星重返所使用的制动火箭包
连续时间间隔	在给定时间间隔内，产品必须满足持续的使用要求	运输飞机
片段型时间间隔	在给定时间间隔内的一个或多个片段内，产品必须满足使用要求	根据要求拍照的侦查卫星

如果时间或用户需求已知，然后要确定所要考虑的产品性能量度，这些量度通常表现为产品输出，还要考虑的是如何用这些性能量度反映产品的整体效能。在前面所介绍的模型中，对效能的度量方法与量化产品能力的方法相同。当然，此量度与应用目标有关，但它通常有多种表述方式，例如本章前面所介绍的通信产品，我们通过定义已传输

的字节数来定义产品性能，并将效能定义为在特定时间段内，产品传输一定量字节的概率。当然也可以把效能定义为预期要传输的字节数。除了把已传输字节作为性能量度，我们还可以用其他量来度量性能，如误码率或传输延迟时间等。

假定性能量度已经确定，我们下面讨论效能量度的两种基本形式：

① 最低性能标准（Minimum Performance Criterion）。最低性能标准具体指定产品输出的量化边界。这些边界定义可接受的性能范围，但在这些边界内对产品的可接受程度（Acceptability）评估是不可取的。因此，这项准则引出了产品性能评估常用的二分法：成功或失败。例如计算机必须在特定时间段内执行一系列（检验计算机能力的）基准问题（Benchmark Problem），或者一枚炸弹的落点必须在距离目标中心 d 英里范围内，这些都称为最低性能标准。

② 整体性能标准（Overall Performance Criterion）。整体性能标准关注与实际输出的完整分布，要在实际的概率分布或相关统计度量（例如期望值）内来考虑。当应用表现或者功能要求比较重要，且最低性能标准在很大程度上是人为确定的时候，可采用此项标准。以计算机基准为例，其他效能量度是给定时间段内的信号输出量或完成基准所消耗的时间。

有些时候，应该采用何种度量方式并不明确。应用目标、产品功能和相关的输出在某些时候规定了评估准则的选择。如果功能的输出是二分的（例如检测产品的输出可能是执行了检测或没有执行检测），那么就要采用最低性能标准，并把成功概率作为量度。对于一个有多种输出或持续输出的产品而言，如果成功边界可以完全人为定义或主要关注点是输出信号的统计特性（例如均值和方差），那么就可以采用整体性能标准。

对于整体性能标准和最低性能标准的选择是非常重要的。在基于最低性能标准的效能量度中，产品输出的具体指定边界决定了可接受的性能范围，并且是划分输出结果成功或失败的标准。对于使用整体性能标准量化产品效能的方法来说，它考虑产品输出的完整分布，并通过适当的统计量（如平均输出）来量化产品效能；对于其输出只返回部分或次要信息的产品而言，可采用整体性能标准来对它的某些性能进行度量。

总之，产品输出和相关的效能量度形成了评估提案设计方案效能的基础。通过建立数学模型来表示这些量度、相关的成本以及实现设计目标所要承担的任务量，可以完成此评估过程。

15.3.3 阶段Ⅲ：建立数学模型

这一阶段将要完成三项任务：

① 选择影响产品效能的变量。

② 定义已有和外加的经济约束。

③ 建立变量间表示效能量度的数学关系。

此处的变量指将在模型中出现的影响产品效能的产品和后勤保障参数。在参数及模型的定义约束内，合理地选择参数值，将会产生一个具有最佳效能的产品。典型的变量会影响效能的主要组成部分：可靠性、维修性、性能和价值。价值可以基于一些共同标准，相互权衡。

这些变量包括复杂性、冗余单元数量、冗余类型、可更换模块的数量和水平、预防性维修的类型和频率等。由于效能分析是一个迭代过程，所以还需要考虑更多的“细节变量”，例如零部件上的应力、维修可达性因素以及测试点的数量和类型等。还应该确定每个变量所带来的负担和收益，例如重量的增加，通过冗余改进可靠性所带来的成本与复杂程度的提升等。把冗余改进可靠性与使用超高可靠性零件带来的负担和收益进行比较，可以完成相关决策。

第二项任务是确定已有和外加的物理、保障和经济约束等。外加物理约束可以明确地定义出来（例如可用的空间大小），也可以通过运行要求引申出来；已有约束由当前技术水平定义，它可能包括可达失效率水平、维护维修率和数据传输速度等。例如资金、产品投放市场的指定日期、可提供维修的人员配置和技术水平这类的约束条件也要列举出来。另外，还需要考虑经济约束和后勤限制，如成本、模拟现场产品所需数据、可用维修人员和技能级别等。还要考虑的经济和后勤约束有：总成本、开发时间、试验产品需求维修人力和技能级别需求等。所有的这些因素必须引入到数学模型中，以确保制造出的产品在特定的约束条件（包括成本、时间计划和保障）下能满足其运行需求。

第三项任务是建立数学模型，它们用来：

① 估计产品效能。

② 评估各种可选方案的有效性。

③ 权衡这些可选方案。

④ 确定产品级别的可靠性和维修性需求。

首先基于产品状态建立一个通用模型。如本章前面所描述的那样，产品状态是由主要产品元素的状态定义的。通过评估这些元素的子功能的表现，可以确定设计和保障决策对产品整体效能的影响。在这个早期阶段，能力分析是一项工程功能，因为直接可用的数据非常有限，所以它更依赖于基本设计原则。完整的分析可以提供概率性能指数，例如一个表示探测概率相对于探测范围的累积分布函数比较适用于评估雷达的性能，而雷达范围方程可以用作工程分析的基础。

此时，还需要进行成本/效能的权衡。例如虽然降低了产品服务寿命的保障成本，但建立高可靠性和维修性却增加了初始投资成本。开发阶段的早期，应该在主要的产品和保障级别进行这种类型的权衡，这样能缩小可选方案的范围。

建立模型并将其应用于可靠性和可维修性相关的成本分析。产品失效会触发保障系统，还要确定某特定产品消耗保障资源的频率。人力和维护时间方面的开支是维修性特征的函数。

成本模型的框架应提供以下信息：

① 主要附属机构运作的成本比较数据。

② 产品单元的保障成本，识别这些单元可能会证明工程变更的合理性。

③ 比较在不同维修级别进行维修的成本信息，与失效频率相关的保障成本的弹性。

④ 在固定投资内，保障所节省的资金和增加的资金和权衡信息，例如新型测试设备的引进。

⑤ 诸如仿真模型和适用数据之类的工具。在开始更改之前，它可以对保障组织内的更改建议进行评估。

更详细的关于成本—效能分析的介绍，参见后续章节。

15.3.4 阶段Ⅳ：获取输入数据

所要输入到模型的数据包括诸如产品零部件的可靠性和维修性参数、成本、重量、所需空间以及与物理、工程或经济因素等相关的数据。最初，这些输入数据可以从以往经验和适当的估计技术中获得。随着早期设计方法变得越来越明确，组件和单元开发的不断推进，应该在此过程中提炼这些输入数据，迭代使用分析模型可以完成提炼。在开发过程中，为了确保效能分析小组能够明确这些数据的生成过程、能够明确需要收集什么数据，我们需要建立分析策略及程序，这一点是非常重要的。

15.3.5 阶段Ⅴ：应用、解释和改进模型

此阶段主要包含以下几个必要步骤：

① 设计一个满足约束条件的产品。

② 计算产品的效能和价值。

③ 将计算所得值与需求相比较。

④ 泛化相应的设计保障因素组合。

⑤ 修订影响因素，重新运行模型。

⑥ 当其他数据可用时，对模型进行优化。

这一过程可以用图15.4表示。实际上，模型只用来评估那些满足物理和经济约束的设计，但是设计的范围通常受客户需求的限制。在一个处于开发中产品的约束已知的情况下，一个概念设计方案可以借助模型转化成实际的硬件配置和能够提供更高效能的保障计划。

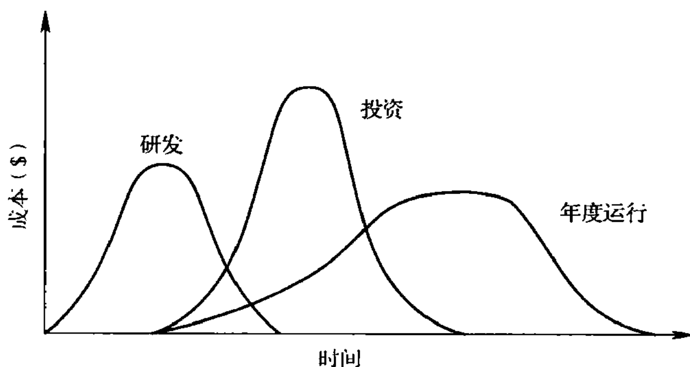


图15.4 模型的实现过程

15.4 成本效能分析

在这一节中，为了描述一种更完整的用于在多个设计方案、运行方法和后勤结构中

做出选择的方法，我们把成本引入到效能图中。我们可以把成本效能分析定义为对比不同方案的过程，此对比过程根据价值收益（效能）和资源消耗（成本）来评估方案满足产品或运行要求的情况。

需要注意的是：度量资源使用情况的量是成本，度量效能的量是价值收益。我们必须认识到，效能可能无法包含一种产品的所有价值元素，成本也可能不会包含所有需要的资源。例如人员技能和计划延误之类的资源需求往往很难转化为成本量度。因此，要确保所有与成本效能分析相关的文档包括那些没有明确定义为成本或效能数值的重要元素。

20 世纪 60 年代初期，成本效能分析在大规模军事开发和采办项目中得到了广泛的应用。它是从几十年前的经济分析工作（也称为“成本效益分析”）中进化而来的，例如 19 世纪 30 年代的洪水控制工程中的经济分析。另外一个相关术语是产品分析，它包含了许多和成本效能分析相同的理念，但它并没有明确定义出最终需要得出的成本和效能数值。为了让决策者可以将此分析与专家决策和直觉相结合起来，从而做出决策，成本效能分析的结果可能会被送至更高级别的系统分析框架中进行分析。

15.4.1　成本分类

根据产品类型和可用数据适用性的不同，产品成本有多种分类方法。成本分类应该重点关注在产品生命周期内消耗的主要资源。基于项目阶段的广泛分类包括与研发、投资和运行相关的成本。表 15.5 给出了每个主要类别的成本类型示例。

① 研发成本：开发产品至生产或采购阶段所需的全部费用。

② 初始投资成本：引入产品到有效库存所产生的全部费用，包括生产或采购费用、设施成本、人员培训、安装和初始备件采购费用等。

③ 运行成本：一旦进入运行库存，产品运行所需的一切费用。虽然研发和投资费用是一次性的，但整个产品生命周期中都会产生运行费用。

图 15.5 中的曲线是这些成本在产品生命周期内的典型分布。生命周期成本（Life-Cycle Cost）通常用来表示在产品的预期寿命内的研发、投资和运行成本的总和。特别要指出的是，在比较可选方案时，要排除所有不影响决策的成本。例如估计几个可选方案的生命周期成本就包括估计其基地级维修的成本。如果所选方案的库房管理费用存在不可预见因素，就要排除这些成本，以简化问题。

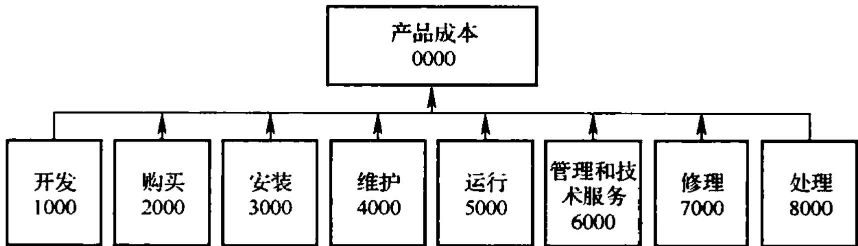


图 15.5　生命周期内产品成本的划分

表 15.5 基于项目阶段的成本分类

	研 发 成 本
设计与开发	初步的调研与设计研究 开发工程及硬件制造 开发使用设备 工业设施
产品试验	试验工具制造 试验工具备件 试验运行 试验支持产品 试验设施 数据采集、筛选、分析和存储 维修、保障、杂项
产品管理和技术指导	初始投资成本
设备	主要应用设备 保障设备 其他设备
库存	应用产品和产品备件 设备保障和零件备件 消耗品
初始训练安装	设施构建 平台修改
杂项投资	技术数据 运输与交通 管理与保障成本
	运 行 成 本
设备和安装替换	主要应用设备 专用设备 其他设备 安装
维修与保障	主要应用设备 专用设备 其他设备
再培训	
库存管理	
管理与技术数据	
设施	
运行成本	人员 燃料 电力 其他

图 15.6 给出了另外一种成本分类方法，这种方法将以上所有八个种类又进行了详细划分。图 15.7 给出了研发成本的细分方法。

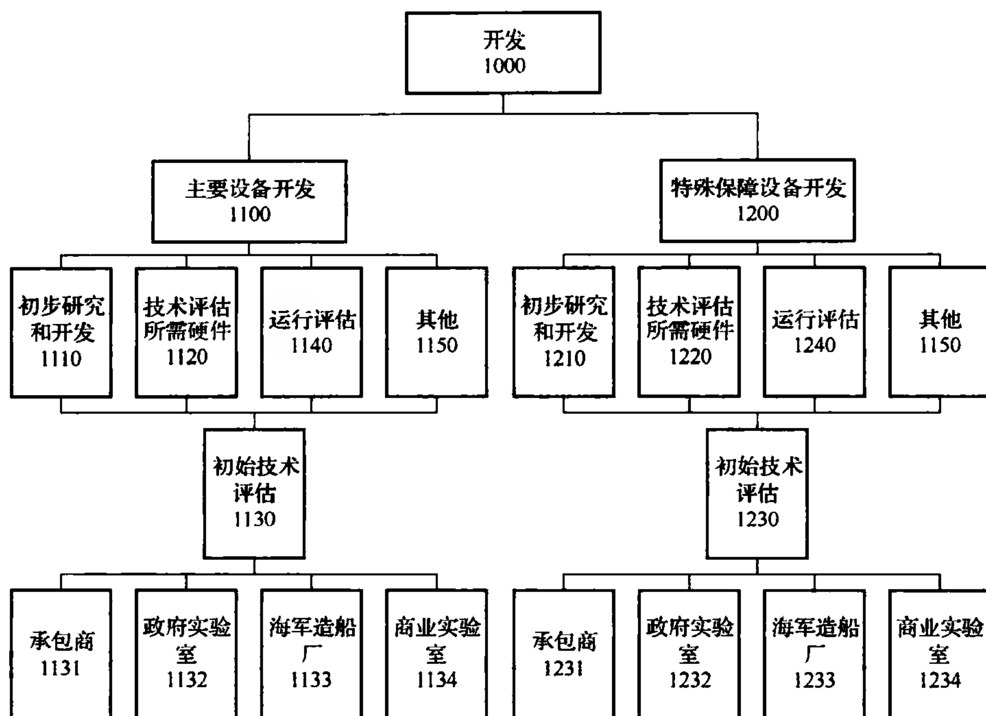


图 15.6　产品总成本（以开发成本为例）

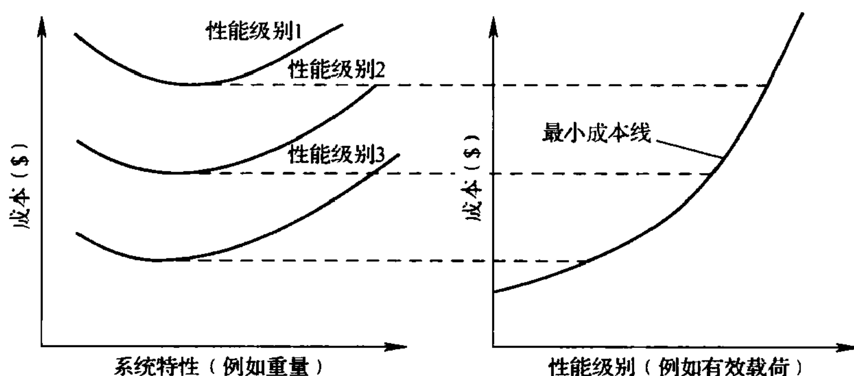


图 15.7　研发成本

15.4.2　成本估计

三种常用的成本估计方法是由下而上（Bottom-Up）、自顶向下（Top-Down）和与相似产品类比（Analogy With Similar Product）。

① 由下而上方法：由下而上方法有时也称为会计法或基层法。它尝试根据费用所包括的元素种类划分费用，并以此估计成本。此划分过程称为任务分解结构（Work Breakdown Structure，WBS）估计内容一般由所需的劳动力和材料成本构成，这些内容

与标准劳动率和材料成本一起，用来估算种类成本；然后将各类费用汇总到更高级别种类，从而建立整体成本估计。更高级别成本（如设施成本）将根据需要引入到整体估计的建立过程之中。

运行和保障（Operation and Support，O&S）成本的估计通常涉及一些由下而上方法，至少在估计后勤和维修成本时需要这类方法。举个简单的例子，如果有200件产品，每件产品每月运行40h，并且每件产品的MTBF为1000运行小时/故障，则

$$\text{预期每月发生故障数} = 200 \times 40/1000 = 8 \tag{15.24}$$

估计人力、材料和其他费用时（例如由于维修设备不在现场所造成的运输成本），需要将产品还原到其运行条件下，以提供对失效成本的估计，然后将所关注的某年或某段时间的费用汇总。除了有效维修成本之外，MTBF和修理时间也决定了需要购买多少备件才能确保产品达到所需的可用性水平。如果应用于某些行为或材料的成本有边界，那么可以把这些成本纳入汇总成本中，以提供对于估计过程中不确定性的度量。有一种简单的方法，即用所有高（低）成本估计可以确定最大（最小）成本的边界。如果已知或可以推断出不确定要素的成本分布，那么就可以用基于随机变量总和的统计程序来建立总成本的分布估计。

② 自顶向下方法：自顶向下方法也称为参数化方法（Parametric Method），它使用历史数据和统计手段，以期寻找高层级成本与研究中适用于产品的一系列参数之间的关系。成本估计关系（Cost-Estimating Relationship，CER）常用于此类成本估算的公式。典型的CER开发适用于一般产品的数据收集（例如雷达产品）。通常，由于影响某一成本种类的因素对其他类别只有很小的影响，甚至会是相反的影响，因此要为每个主要成本种类（例如研发、投资、运行）确定单独的CER。例如一个生产极高可靠性产品项目的开发成本会超过生产具有典型可靠性水平产品项目的开发成本，但开发者所期望的是通过减少失效次数而大大减少产品的运行成本。

需要考虑的数据包括相关产品的性能因素、物理属性和成本等，另外还要收集与技术因素和采办环境因素相关的数据。例如在计算机成本估计中，了解每台计算机大规模集成电路的数量可能就很重要。如果某些产品只有单一客户源采购，而另外一些则存在多家竞争采购，这会对支付价格产生重要的影响。可靠性需求或保修时间等因素也会影响成本。

使用诸如多重回归分析（Multiple Regression Analysis）之类的技术，可以把与所关注成本高度相关的因素作为“显著因素”筛选出来，并将其应用到预测公式中。对于雷达产品而言，CER可能取决于覆盖范围和灵敏度等技术因素；对于计算机产品来说，CPU速度和内存容量可能是很好的成本预测指标；对于飞机来说，飞行范围、速度和承载能力显然是成本预测的重要考虑因素。用于建立CER的统计过程还提供度量关系强度的量，并且能通过诸如置信区间的技术度量估计中的不确定因素。

在决定将哪些因素包括进数据库、在非正常事件发生时应对成本数量做出如何调整、在决定应该用什么筛选标准确保最终公式推导出的工程结果与统计结果同样有效时，都需要制定大量的工程决策。

为了阐明 CER，航空无线电设备公司（ARINC）在几年前开发了用来预测雷达产品开发成本的公式：

$$\ln C_{st}=0.784+0.205\ln A_{pr}+0.165D_{dev}+0.151\ln P_k+0.028S+0.082SC+1.37TD \tag{15.25}$$

式中 A_{pr}——天线孔径（in^2）；

D_{dev}——开发程度（评分：0、1、2）；

P_k——峰值功率（kW）；

S——灵敏度（-dBm）；

SC——专用电路数量；

TD——开发类别（新类型=2，改进类型=1）；

C_{st}——开发成本与第一个样机成本。

③ 与相似产品类比法：类比方法指使用具有类似特性产品的成本数据，然后在现有产品和评估产品不同的情况下，对这些数据进行修改。数据可以取自价格表和以前采购合同中所列出的费用。在很多情况下，数据调整将涉及外推法（Extrapolation），例如增加计算机 CPU 速度、增加发动机的功率范围并降低燃料消耗或为了增加软件购买量而提供的图像用户界面等。同样，此处也需要经验和良好的工程决策来决定哪个历史数据具有关联性，并根据当前产品和以往产品的不同来进行数据调整。当类比产品与目标产品本身具有相似性且具有相似的数据采集环境时，类比方法显然会提供较好的分析结果。随着产品差异的增加，类比估计的准确性将降低，且估计得不确定程度很难量化。

15.4.3 成本调整

为了考虑初步估计过程中所忽略掉的巨大成本影响，需要在成本估计中采用一些“标准”技术。其中，两种技术是规模经济（Economy of Scale）和折扣。

① 规模经济（Economy of Scale）：通常，制造过程中的产品产量将对单位成本产生重大影响。这就是规模经济现象，它可以用多个因素来进行解释（例如购买大量原材料的能力，将固定成本分散到大量产品和学习效应等）。这种现象可以用一个广义方程来表达：

$$K_{CA}=(P^*/P)^{a_c} \tag{15.26}$$

式中 P^*——标准生产量；

P——计划生产批量；

a_c——常数（正数）；

K_{CA}——根据生产批量 P^*，来调整成本的系数。

一种典型的学习曲线公式反映了随着使用率或应用率的增加（也就是生产、维护），相应的时间减少量（或者是其他资源消耗量，如工时）的情况，即

$$T(R_c)=A_1R_c^{b_c} \tag{15.27}$$

式中 $T(R)$——第 R 个产品单元所需要的时间；

A_1——第 1 个产品单元所需要的时间；

R_c——累积单元数；

b_c——常数（负数）。

如果我们假定资源利用率随使用率的双倍增加呈固定百分比下降趋势，那么常数 b_c 可以通过式（15.28）来确定：

$$b_c = [\ln(percent) - 2]/\ln(2) \tag{15.28}$$

② 折扣：如果工作需要执行一段时间，那么最好要求预先支付费用，而不要等到工作完成后再要求支付费用，这样可以避免产生支付担保等相关的问题。通过预先支付，你可以先将得到的钱存放在银行中。这样的话，在工作结束后，你得到的将不仅仅是合同上的金额，还能够从银行赚取利息。但付款人也会意识到这一点，如果他提前支付，他也将失去这些潜在利息，因为它必须先从银行提钱来付款。因此，他可能会建议支付折扣金额，以弥补其利息损失。如果折扣率与银行存款的利率相同，那么收款人会在工作结束后获得全额的款项，同时，付款人银行账户的金额也将和他在任务结束时的付款额保持一致。

因此，折扣是一个考虑货币时间价值的过程。应该将这个概念应用于所有年度支出，这样，每年的费用都将会是一个常数值。这对于做出准确评估是很有必要的。年度支出的折扣量称为现值（Present Value），计算方法如下：

$$\mathrm{PV} = C_{fe}/(1 + i_d)^n \tag{15.29}$$

式中 PV——现值；

n——未来周期数；

C_{fe}——未来 n 个周期内的支出；

i_d——每周期的折扣率。

表15.6介绍了一个两种产品的成本比较的简单示例（折扣率为10%）。这两种产品在折扣前的总支出都为3500美元，但因为产品 B 的大多数支出发生在最后2年中，相比70%支出发生在前3年中的产品 A，它的总折扣成本较低。所以在其他条件都相同的情况下，B 是成本较低的产品。

表15.6 两种产品的成本比较

年	折扣因子	产品 A 的项目支出	产品 A 的项目折扣成本	产品 B 的项目支出	产品 B 的项目折扣成本
1	0.91	1000	910	500	455
2	0.83	500	415	500	415
3	0.75	1000	750	500	375
4	0.68	500	340	1000	680
5	0.62	500	310	1000	320
6		3500	2725	3500	2545

要注意的是，折扣和通货膨胀是两个不同的概念。通货膨胀指单位货币和基准年相比之下的购买能力，折扣指持有货币的价值。成本效能分析必须考虑货币的时间价值，但除非存在特殊情况，通常不需要考虑通货膨胀。

15.4.4 成本的不确定性和敏感性

在讨论可用的成本估计类别的过程中，我们曾表明，成本的不确定性是一个必须考虑的问题。无论是用由下而上的方法对一小部分零件或者小范围活动的成本进行估计，还是在自顶向下的方法中应用 CER 的结果，又或是用相似产品类比法来调整成本，不确定性都是常见的问题。

CER 方法为它所采用的统计技术提供了直接处理不确定性的方式。CER 通常用标准偏差或置信区间因子的形式度量不确定性。但在某些情况下，还存在与应用到产品的历史数据的可用性相关的不确定性、与环境相关的不确定性，这些都需要予以考虑。在对这些不确定性进行预测时，应该使用外推法（Extrapolation）而不是内插法（Interpolation）。对于另外两种方法，虽然我们在前面提到过，它们较难给出量化不确定性的量度，但在由下而上的方法中，可以使用统计理论对不确定成本数求和，就如同我们进行项目评估和技术审查（Program Evaluation and Review Technique，PERT）分析时的做法一样。此处，很有可能要根据悲观结果和乐观结果数来建立项目时间的分布。对于所有三种方法，当涉及对大规模产品的生命周期成本进行估计时，需要确定所有相关以及重大的成本元素。

不管是否能提供不确定性的量化指标，成本分析师的职责都是提供尽可能明确的不确定性信息，并帮助决策者评价不确定性产生的影响。敏感度分析（Sensitivity Analysis）是深入研究分析不确定性所产生影响的方法之一。改变一个不确定的成本变量——如人力工资率——分析师可以确定总成本对于人力工资率改变的敏感程度。如果两产品具有相同的效能，但使用元素成本的最佳估计得出的结果是 *A* 比 *B* 的成本低，那么此结果表明了相对于人力工资率的偏差，选择 *A* 的敏感程度要高；当人力工资率极其有利于 *B* 时，如果仍做出同样的决定，那么人力工资率的不确定性将不再那么重要。

15.4.5 综合考虑效能和成本

我们已经建立了一些用来制定评估可选产品效能和成本量度的方法。如果有 *A* 和 *B* 两个产品，当参考表 15.7 对效能和成本进行比较的时候，我们可以得到四种可能性。

表 15.7 产品 *A* 和 *B* 的效能成本比较

	效 能	成 本	决 策
情况 1	*A* 更好	*A* 更好	*A*
情况 2	*A* 更好	*B* 更好	?
情况 3	*B* 更好	*A* 更好	?
情况 4	*B* 更好	*B* 更好	*B*

情况 1 和 4 表明某产品比另外一产品占有绝对优势，而情况 2 和 3 的比较结果就不是很明显。以情况 2 为例，如果 *A* 在效能方面比 *B* 好很多，但在成本方面却和 *B* 几乎相等，那么就很有可能选 *A*。但当把成本作为一个主要衡量标准，情况又会如何？即便某产品占绝对性的优势，但在与另外一产品进行比较时，其成本估计中的不可量化因素和不确定性可能会使决策结果变得没有看上去那么明显。

另外一个难题是“杠杆（Leverage）作用”或成本、效能模型中没有明确考虑的因素对产品产生的影响。例如表15.8所列出的用于新型直升机的两种发动机的相关参数。表面上看来，由于发动机A比B在成本上节省了1500万美元，所以理论上应该选择发动机A。但是，假如发动机B的设计允许对其进行修改，经修改后它还可以在飞机上使用，情况又如何？如果选择发动机B，飞机发动机的预期开发成本将减少3000万美元，那么考虑到这种“杠杆作用”，发动机B将是更好的选择。

表15.8　杠杆作用

	动力发动机A	动力发动机B
总成本（$）	60 000 000	75 000 000
效能	0.95	0.95

虽然我们无法为可选设计方案或产品的成本和效能值提供一个统一的决策方法，但通常而言，决策者知道成本和效能值总比不知道要好。现在我们将讨论一些在决策过程中对决策者有所帮助的技术。

产品设计研究：我们将首先讨论使用效能分析为给定产品类型选择设计方案的过程。一个需要用到的分析输出是成本—性能曲线（Cost Versus Effectiveness Trade-Off Curve）。以某新型运输机为例，我们所要关注的是特殊性能参数（如载荷），并要分析某些产品特征（如发动机推力或飞机重量）。图15.8在几个不同性能级别上描绘了成本和产品特征之间的关系。在经济学中，这些曲线被称为等量曲线（Isoquant）。如果我们为每个性能级别取最小成本并画出成本—性能曲线，就可以得出所有性能级别上最低成本的权衡曲线，如图15.9所示。在决策者选择设计方案时，此权衡曲线是非常有用的工具。实际上，如果成本和性能量度包括了所有相关因素，此曲线会提供一个优化的解决方案。

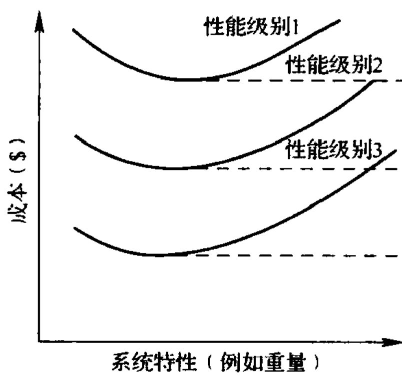


图15.8　等效性能曲线

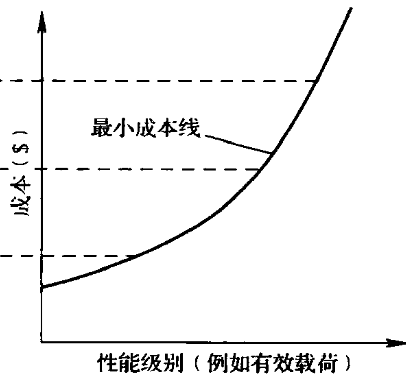


图15.9　成本—性能曲线

产品比较研究：此项研究比较两个或更多产品对同一应用目标的完成情况。所比较产品在设计或运行方面也许有相似之处，也许没有（例如载货汽车或火车都可以用于运输材料，并且这两种运输方式都能满足目标要求）。如果能为每个产品建立一个如前

所述的成本—性能曲线，那么我们就可以得到类似于图 15.10 的分析结果。此处我们可以看到，对于效能值小于 E_0 的情况，产品 A 将提供低成本的解决方案；但对于较高效能值的情况，产品 B 则是较好的解决方案。有两种可用方式处理这种情况：一种是固定效能，另外一种是固定成本。

① 固定效能：如果可以具体指定所需的效能或性能级别，那么在不考虑任何“杠杆作用”的情况下，以最低的成本满足要求的产品将是首选项。

② 固定成本：如果总成本已被编入预算，那么在不考虑任何“杠杆作用”的情况下，在该成本价位上能够提供最大效能的产品将是首选项。

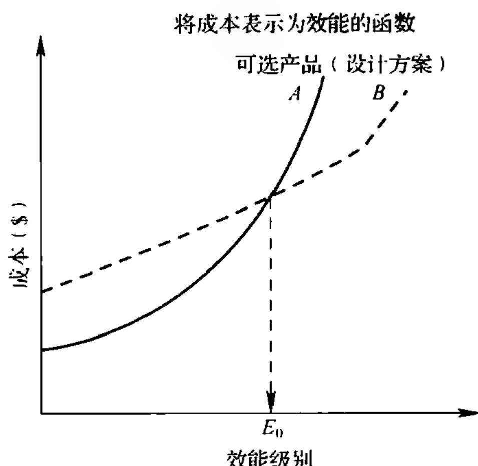


图 15.10　产品 A 和 B 的成本效能曲线

如果可行的话，固定效能或固定成本将会是个比较理想的方式。在许多情况下，固定效能或固定成本是不可行的，但它们可能会被限制在一定的范围内。例如成本效益分析师面临的问题是向决策者提供什么样的信息，使其可以选择最佳的产品。满足效能至少为 E' 且总成本不超过 C'，图 15.11 所示的阴影区域说明了这种情况。在该区域内，成本与效能的任意组合都表示一个可接受的解决方案。从所示的例子中，我们可以看到 A 和 B 都不占主导地位，选择何种方案仍不是很明确。

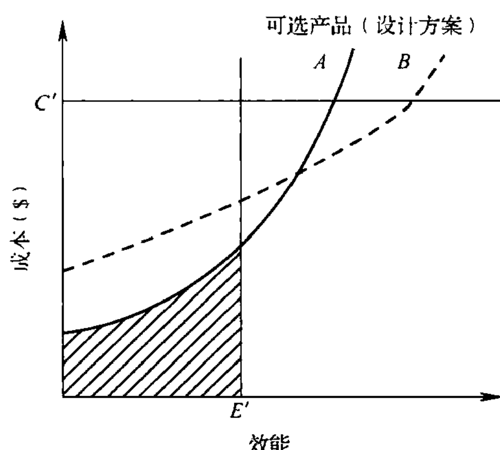


图 15.11　成本效能曲线：接受区域

一种常用的方法是计算效能对成本的比率，所得到的结果值是最划算的。虽然在某些情况下可以使用这种方法，但已有人对这种方法提出了批评，认为它是“片面的”。分析标准的缺失并不会降低成本效能分析的价值，这意味着必须向决策者提供尽可能多的信息。虽然这些信息不关注某个单独的数据，但我们可以用一种便于将其与决策者的专家决策相结合的方式显示出来。这将要求成本效能分析师对产品进行灵活建模，从而使得评估能适应于不断变化的信息需求。

15.5　总结

这一章节介绍了 Markov 模型框架，它将产品的可靠性、维修性和性能特征结合到了整体效能中。虽然这个模型框架拥有很多简化的特征，但正如第 15.2.2 节所描述的

通信系统示例一样，经过适当扩展后，可以把该模型框架作为复杂产品分析的基础。本章特别关注于成本问题，讨论了在可选设计方案的效能评估过程中可用的分析方法和所要考虑的因素；还回顾了三个主要成本元素——研发、初始投资和运行以及成本的估计方法；此外还讨论了规模经济、折扣和成本的不确定性和灵敏度；最后一部分讨论了决策制定者在综合考虑效能和成本过程中所面临的困难。

参考文献

Shooman, M. L. 1990. Probabilistic reliability: An engineering approach. Malabar, FL: Robert E. Krieger.

辅助阅读材料

ARINC Research Corporation. 1969. Guidebook for systems analysis/cost effectiveness. U. S. Army Electronics Command.

DARCOM P700-6 (Army), NAVMAT P5242 (Navy), FLCP/AFSCP 800-19 (Air Force). 1977. Joint-design-to-cost guide: Life-cycle cost as a design parameter. Washington, DC: U. S. Department of Defense.

Dhillon, B. S. 1989. Life-cycle costing: Techniques, models and applications. New York: Gordon and Breach, Science.

English, J. M. 1968. Cost effectiveness—The economic evaluation of engineering systems. New York: John Wiley & Sons.

Fabrycky, W. J., and B. S. Blanchard. 1991. Life-cycle cost and economic analysis. Englewood Cliffs, NJ: Prentice Hall.

Goldman, T., ed. 1967. Cost-effectiveness analysis. New York: Frederick A. Praeger

Michaels, J. V., and W. P. Wood. 1989. Design to cost. New York: John Wiley & Sons.

Ostwald, P. F. 1992. Engineering cost estimating, 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Quade E. S., and W. I. Boucher. 1968. Systems analysis and policy planning. New York: American Elsevier.

Taguchi, G. 1992. Taguchi methods: Research and development. Englewood, CO: ASI Press.

第16章　工艺能力与过程控制

16.1　引言

质量是产品满足工艺标准的能力。本章将介绍工艺能力（Process Capability）的概念，还将介绍用来实现并保持零件和产品质量的统计过程控制技术（Statistical Process Control，SPC）的基本知识。

16.2　平均检出质量

平均检出质量（Average Outgoing Quality，AOQ）是衡量产品质量的一个指标。它通常被定义为在最终质量控制检查中所检出的超规格产品和产品总数的比例，通常以百万分之几（ppm）为单位［Ackermann and Fabia，1993］。AOQ越高，表明越多的缺陷数和越糟糕的质量水平。

$$\text{AOQ} = \frac{\text{过程曲线下方阴影部分面积}}{\text{过程曲线下方总面积}} \times 10^6 \tag{16.1}$$

其中，USL是产品规格上限；LSL是产品规格下限；μ是过程均值。

例如某制造商通过外观、机械和电气测试去衡量某电子产品的AOQ。外观和机械测试包括尺寸、可焊度和弯角检查，电气测试包括室温、高温和低温下的功能和参数测试。平均检出质量AOQ由公式（16.1）确定，如图16.1所示。

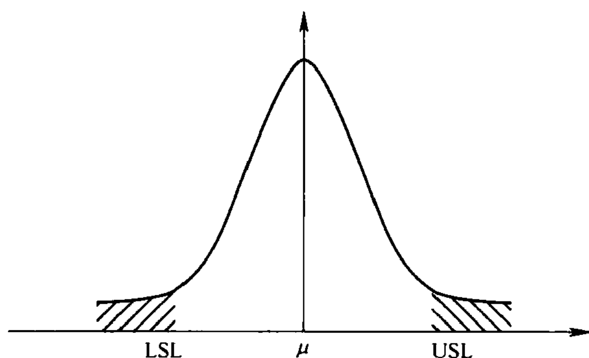


图16.1　AOQ示意图

在不同的制造商之间，AOQ的计算公式可能会有所差异，例如基于JEDEC标准的JESD 16-A［JEDC，1995］就使用以下定义公式：

$$\begin{aligned}\text{AOQ} &= P \times \text{LAR} \times 10^6 = \frac{D}{N} \times \text{LAR} \times 10^6 \\ &= \frac{D}{N} \times \frac{\text{AL}}{\text{TL}} \times 10^6\end{aligned} \tag{16.2}$$

式中，D——不合格品总数；

N——参与测试产品总数；

LAR——批次接收率；

AL——接收批次总数；

TL——参与测试批次总数。

一个名叫 IDT 的半导体制造商定义了下面的 AOQ 公式：

$$AOQ = P \times 10^6 = \frac{D}{N} \times 10^6 \tag{16.3}$$

式中，D 表示不合格品总数；N 是参与测试产品总数。大多数制造商都是基于不合格品数和样本总数来计算其 AOQ。

16.3 工艺能力

AOQ 是产品离开生产设备时衡量其质量水平的一个参数；而工艺能力（Process Capability）是评价产品满足客户要求能力的量，通常在关键步骤中进行度量。工艺能力评估是确定一个过程在其固有波动限制下，满足产品要求和规格的能力。它有助于识别过程所发生的变化，并能确定不能满足需求的产品或服务的百分比。如果过程没有足够的能力使产品符合要求，那么就必须选择其他过程；在某些情况下，如果规范的设定不切实际，就需要更改产品规范。

图 16.2 显示了一个产品规格界限，这些界限通常只根据客户的要求制定，并不能反映一个过程的能力。规格界限可用来确定产品是否满足客户的要求。图 16.2 是一个用正态分布曲线表达的规格界限。

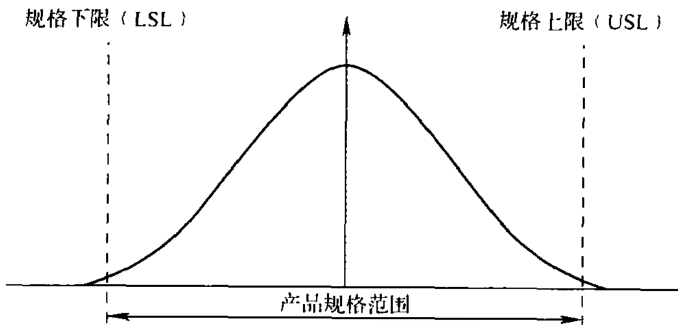


图 16.2　产品规格界限与客户需求符合度的度量

确定工艺能力，第一步就是确定过程的总均值$\bar{\bar{X}}$和平均极差$\bar{R}$，然后确定 USL 和 LSL。利用下面的公式，根据所用的控制图来计算过程的标准差：

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \text{或} \ \hat{\sigma} = \frac{\bar{s}}{c_4} \tag{16.4}$$

其中，$\bar{R}$和 $\bar{s}$ 是一个过程在可控时间内每个子组的极差和标准差；d_2 和 c_4 是常数，它们是根据子组样本量确定的。过程均值可以用$\bar{\bar{X}}$、$\bar{X}$和$\bar{X}$来估计。

比较一个工艺的容差宽度与 6 倍标准偏差的区别，可以表示其稳定程度。比较工艺的 6 倍标准偏差与订货规范（Customer Specification）的偏差，可以得到工艺能力值。评估工

艺能力的参数包括 C_p、C_r（C_p 的倒数）、C_{pl}、C_{pu}和 C_{pk}。C_p 可用公式（16.5）进行计算：

$$C_p=\frac{\text{USL}-\text{LSL}}{6\hat{\sigma}} \tag{16.5}$$

C_p 可以利用正态概率分布曲线预计新产品的不合格率。当 $C_p<1$ 时，工艺变异超过了规格界限，正在生产不合格品；当 $C_p=1$ 时，工艺刚好满足规格，但至少还会产出 0.3% 的不合格品，如果工艺中心不等于规格中心的话，不合格率会更高；当 $C_p>1$ 时，工艺变异小于规格界限；但是，假如工艺中心不等于规格中心的话，仍然有可能产生不合格品。C_p 值和规格界限的这三种关系如图 16.3 所示。

工艺能力指数 C_{pl}、C_{pu}（单边规格界限）和 C_{pk}（双边规格界限）不仅考虑了工艺的变异情况，还考虑了工艺均值的位置。因此，工艺能力是综合描述过程曲线偏离规格中心的程度和过程变差波动范围的程度。C_{pk}是用来描述工艺能力的一个指数，而且必然是 C_{pl}和 C_{pu}中的较小值。假如过程输出近似正态分布并处在统计受控状态，C_{pk}可以用来估计不合格率的期望值：

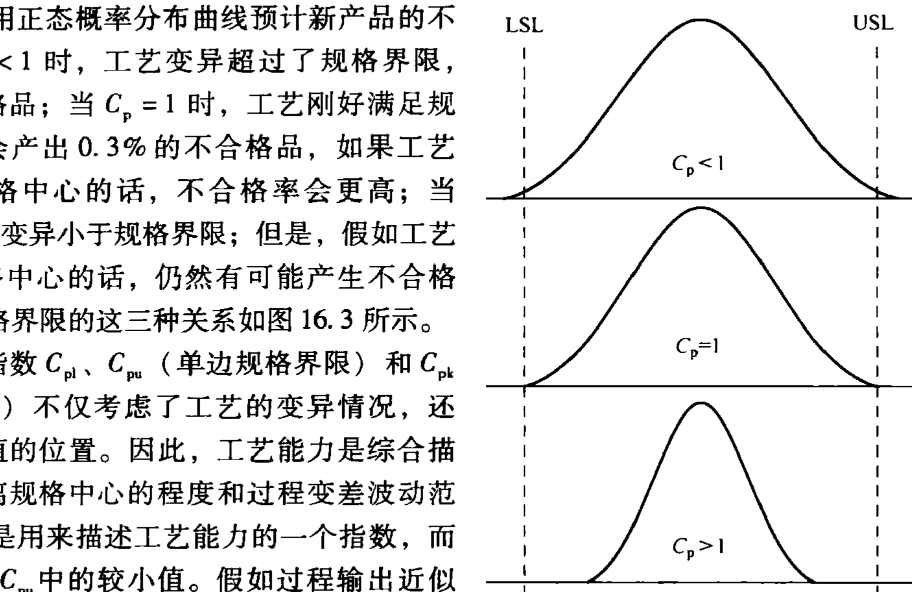


图 16.3　C_p——简单的工艺能力

$$C_{pl}=\frac{\bar{\bar{X}}-\text{LSL}}{3\hat{\sigma}},C_{pu}=C_{pu}=\frac{\text{USL}-\bar{\bar{X}}}{3\hat{\sigma}} \tag{16.6}$$

$$C_{pk}=\min\{C_{pu},C_{pl}\} \tag{16.7}$$

图 16.4 是一个部分无法满足需求的工艺。图中过程的 $C_p>1$，但是其能力却由于工艺中心不在规格中心而下降。

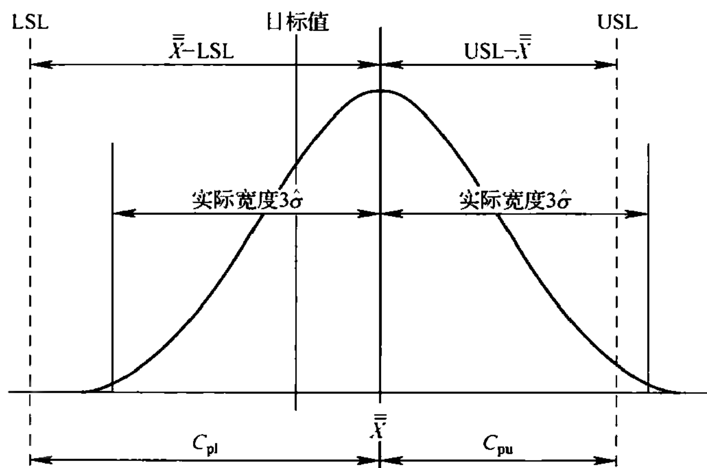


图 16.4　部分无法满足需求的工艺

如果工艺不能持续地生产出合格的产品，就必须确认和纠正引起工艺波动的一般原

因。常见的措施包括安排使用其他机器、购买一台新设备、加强员工培训以减少操作偏差、要求供应商执行统计过程控制等。

案例 16.1

从冲压裁剪工艺的控制图中获得以下统计量：$\overline{\overline{X}}=212.5$，$\overline{R}=1.2$，$n=5$。这项工艺输出特性的规格范围是 210 ± 3，也就是说，上限 USL = 213，下限 LSL = 207。计算 C_{p} 和 C_{pk} 以及相应的不合格率。

解：

$$\hat{\sigma}=\frac{\overline{R}}{d_2}=\frac{1.2}{2.326}=0.516$$

$$C_{\mathrm{p}}=\frac{\mathrm{USL}-\mathrm{LSL}}{6\hat{\sigma}}=\frac{213-207}{6(0.516)}=\frac{6}{3.096}=1.938$$

$$C_{\mathrm{pl}}=\frac{\overline{\overline{X}}-\mathrm{LSL}}{3\hat{\sigma}}=\frac{212.5-207}{3(0.516)}=\frac{5.5}{1.548}=3.553$$

$$C_{\mathrm{pu}}=\frac{\mathrm{USL}-\overline{\overline{X}}}{3\hat{\sigma}}=\frac{213-212.5}{3(0.516)}=\frac{0.5}{1.548}=0.323$$

$$C_{\mathrm{pk}}=\min\{C_{\mathrm{pl}},C_{\mathrm{pu}}\}=0.323$$

由于 $C_{\mathrm{pk}}<1$，这个工艺正在制造不合格品。图 16.5 也直观地说明了这个问题。

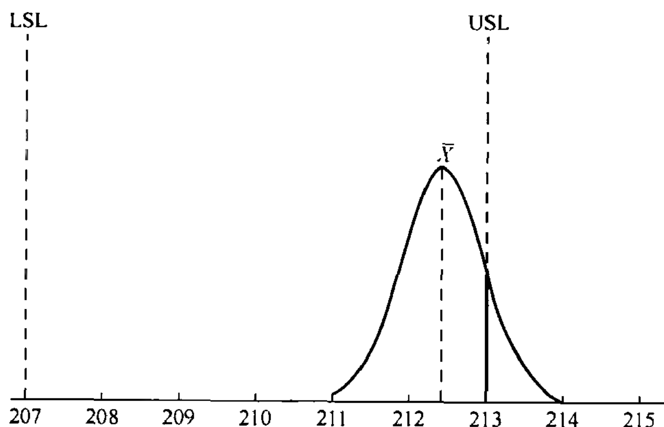


图 16.5　能力不足的工艺

不合格品率的计算：假设工艺近似正态分布并且是统计受控的，C_{pk} 的计算过程也能用来估计不合格品率。图 16.5 中超出规格的阴影区域面积可以用来确定不合格品率。为了确定阴影面积，必须先计算下面的因子：

$$z_1=\frac{\mathrm{LSL}-\overline{\overline{X}}}{\hat{\sigma}}=\frac{207-212.5}{0.516}=-10.68$$

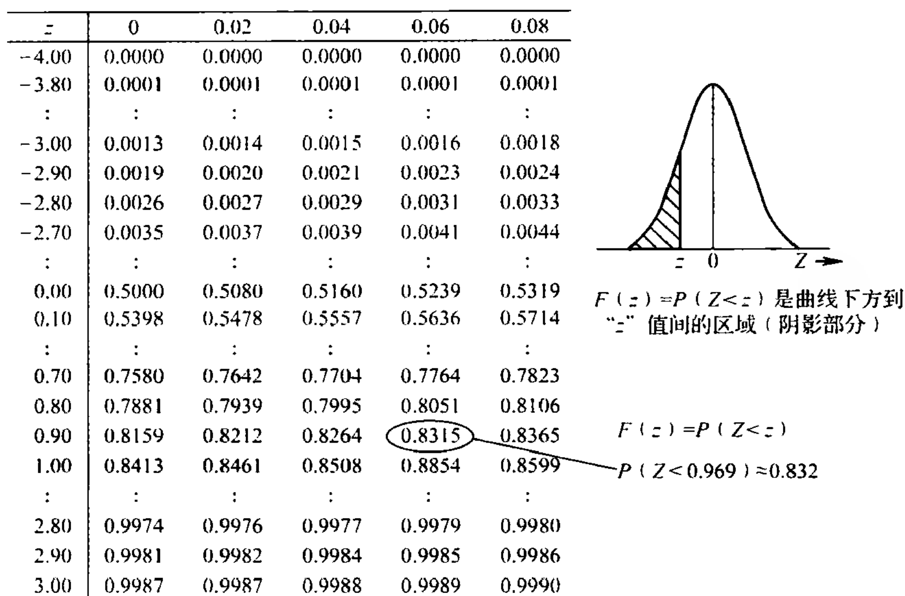
$$z_2=\frac{\mathrm{USL}-\overline{\overline{X}}}{\hat{\sigma}}=\frac{213-212.5}{0.516}=0.969$$

$Z>\mathrm{LSL}$ 的不合格品率 $=F(z_1)$；其中，$F(z_1)=0$（约等于）。

$Z>$ USL 的不合格品率 $=[1-F(z_2)]$；其中，$1-F(z_2)=1-0.832=0.168$。

$-F(z)=P(Z<z)$ 是标准正态分布中 z 的累积概率，见图 16.6。

总不合格品率 $=F(z_1)+[1-F(z_2)]=16.8\%$。



z	0	0.02	0.04	0.06	0.08
-4.00	0.0000	0.0000	0.0000	0.0000	0.0000
-3.80	0.0001	0.0001	0.0001	0.0001	0.0001
:	:	:	:	:	:
-3.00	0.0013	0.0014	0.0015	0.0016	0.0018
-2.90	0.0019	0.0020	0.0021	0.0023	0.0024
-2.80	0.0026	0.0027	0.0029	0.0031	0.0033
-2.70	0.0035	0.0037	0.0039	0.0041	0.0044
:	:	:	:	:	:
0.00	0.5000	0.5080	0.5160	0.5239	0.5319
0.10	0.5398	0.5478	0.5557	0.5636	0.5714
:	:	:	:	:	:
0.70	0.7580	0.7642	0.7704	0.7764	0.7823
0.80	0.7881	0.7939	0.7995	0.8051	0.8106
0.90	0.8159	0.8212	0.8264	0.8315	0.8365
1.00	0.8413	0.8461	0.8508	0.8854	0.8599
:	:	:	:	:	:
2.80	0.9974	0.9976	0.9977	0.9979	0.9980
2.90	0.9981	0.9982	0.9984	0.9985	0.9986
3.00	0.9987	0.9987	0.9988	0.9989	0.9990

图 16.6　抽样累积正态分布表及不合格率估计

16.4　统计过程控制

统计过程控制（SPC）是通过测量样本均值变化趋势和散布范围来监控过程质量特性的一种技术，它取代了生产完成后对最终产品的检查。

16.4.1　控制图：确认变异来源

控制图（Control Chart）是趋势图的一种类型，它的上、下限可以通过统计的方法得出，用来确定过程是否“受控”。当一个过程的变异随机且可预测时，我们认为它是受控的。控制图用来评价过程变异及其来源，并对其进行监控和持续改进。

随机变异是由过程内部各步骤的相互影响而引起的。当过程超出控制界限时，可查明的变异可能就是造成这个现象的原因。它是由一些特殊原因造成的。

控制图有助于确定在过程中发生变异的类型。通过使用控制图，我们可以从过程波动的一般原因中分辨出特殊原因。控制图还能作为一种控制工具，帮助改进过程，提高过程的一致性和可预测性。

16.4.2　构建控制图

控制图有很多种类型，控制图是否合适由数据类型决定。图 16.7 介绍了不同类型的数据及其相应的控制图；图 16.8 给出了如何根据图 16.7 中的信息选择控制图的准则。按照图 16.9 给出的步骤，可以构建控制图。为了计算合适的统计数据，需要了解

不同的方法及其常数。表16.1和表16.2分别给出了一些公式和常数，它们分别用于构建计量型数据（Variable Data）和计数型数据（Attribute Data）控制图。表16.3和表16.4给出了公式中所用到的常数。

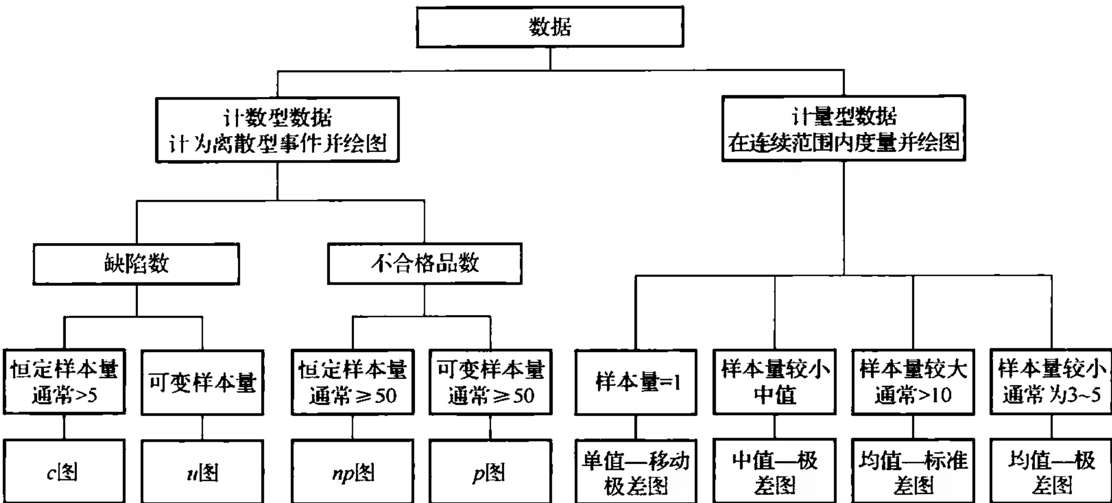


图16.7 控制图选择过程

表16.1 计量型数据的控制图所用的常数和公式

控制图类型	样 本 量	中心线①	控 制 界 限
均值和极差 均值—极差图	$n<10$，通常为3~5	$\bar{\bar{X}}=\dfrac{(\bar{X}_1+\bar{X}_2+\cdots+\bar{X}_k)}{k}$ $\bar{R}=\dfrac{(R_1+R_2+\cdots+R_k)}{k}$	$UCL_{\bar{X}}=\bar{\bar{X}}+A_2\bar{R}$ $LCL_{\bar{X}}=\bar{\bar{X}}-A_2\bar{R}$ $UCL_R=D_4\bar{R}$ $LCL_R=D_3\bar{R}$
均值和标准差 均值—标准差图	n通常≥10	$\bar{\bar{X}}=\dfrac{(\bar{X}_1+\bar{X}_2+\cdots+\bar{X}_k)}{k}$ $\bar{S}=\dfrac{(S_1+S_2+\cdots+S_k)}{k}$	$UCL_{\bar{X}}=\bar{\bar{X}}+A_3\bar{S}$ $LCL_{\bar{X}}=\bar{\bar{X}}-A_3\bar{S}$ $UCL_S=B_4\bar{S}$ $LCL_S=B_3\bar{S}$
中值和极差 中值和极差图	$n<10$，通常为3~5	$\bar{X}=\dfrac{(\bar{X}_1+\bar{X}_2+\cdots+\bar{X}_k)}{k}$ $\bar{R}=\dfrac{(R_1+R_2+\cdots+R_k)}{k}$	$UCL_X=\bar{X}+A_2\bar{R}$ $LCL_X=\bar{X}-A_2\bar{R}$ $UCL_R=D_4\bar{R}$ $LCL_R=D_3\bar{R}$
单值和移动极差 单值和移动极差图	$n=1$	$\bar{X}=\dfrac{(X_1+X_2+\cdots+X_k)}{k}$ $R_m=\lvert(X_{i-1}-X_i)\rvert$ $\bar{R}_m=\dfrac{(R_1+R_2+\cdots+R_k)}{k-1}$	$UCL_X=\bar{X}+E_2\bar{R}_m$ $LCL_X=\bar{X}-E_2\bar{R}_m$ $UCL_{R_m}=D_4\bar{R}_m$ $LCL_{R_m}=D_3\bar{R}_m$

① k=子组数；X=每个子组的中值；$\bar{X}=\dfrac{\sum_{i=1}^{n}X_i}{n}$。

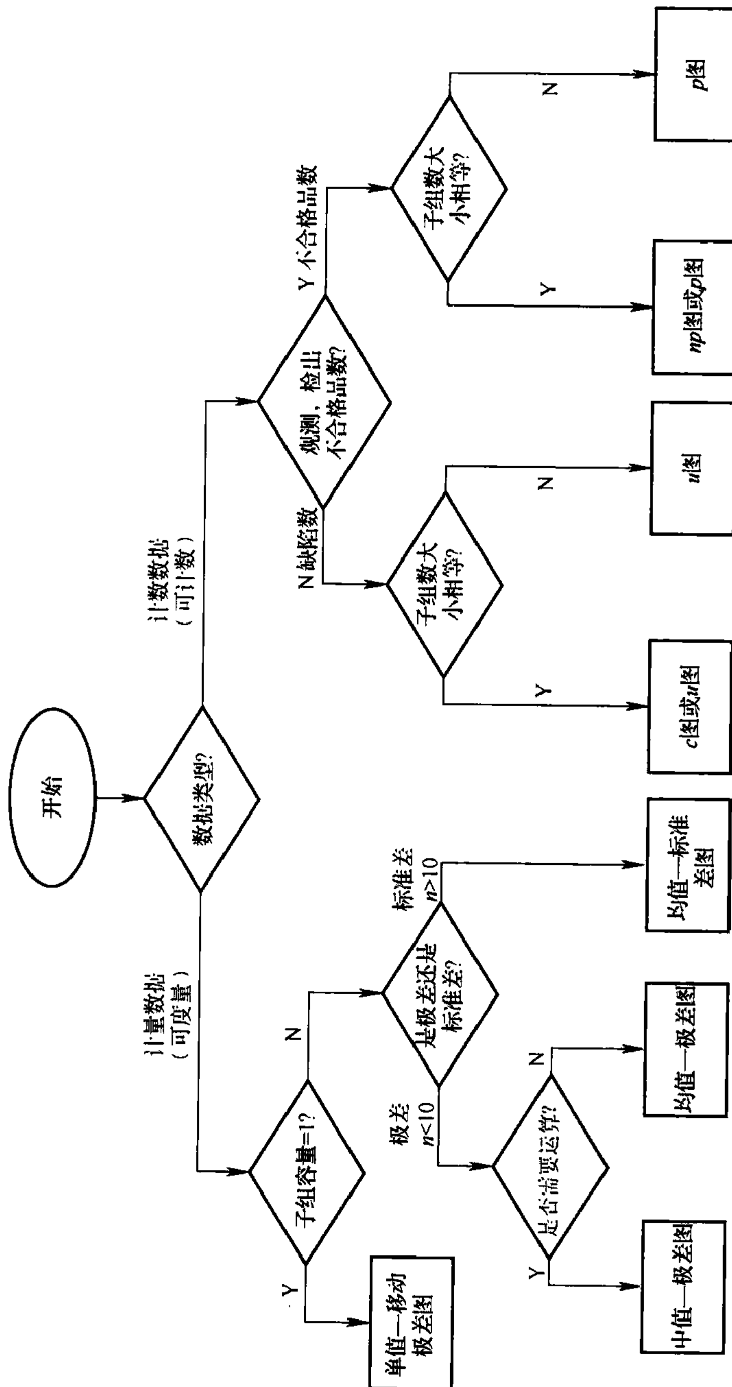


图 16.8 选择控制图的准则

表16.2 计数型数据的控制图所用的常数和公式①

控制图类型	样本量	中心线	控制界限②,③
不合格率 p 图	可变，通常为 $n>50$ 或 $n=50$	对于每个子组： $p=np/n$ 对于所有子组： $\bar{p}=\dfrac{\sum np}{\sum n}$	$UCL_p=\bar{p}+3\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}$ ② $LCL_p=\bar{p}-3\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}$ ②
不合格品数 np 图	恒定，通常为 $n>50$ 或 $n=50$	对于每个子组： np = 不合格品数 对于所有子组： $n\bar{p}=\dfrac{\sum np}{k}$	$UCL_{np}=n\bar{p}+3\sqrt{n\bar{p}(1-\bar{p})}$ $LCL_{np}=n\bar{p}-3\sqrt{n\bar{p}(1-\bar{p})}$
缺陷数 c 图	恒定	对于每个子组： c = 缺陷数 对于所有子组： $\bar{c}=\dfrac{\sum c}{k}$	$UCL_c=\bar{c}+3\sqrt{\bar{c}}$ $LCL_c=\bar{c}-3\sqrt{\bar{c}}$
单位缺陷数 u 图	可变	对于每个子组：$u=c/n$ 对于所有子组： $\bar{u}=\dfrac{\sum c}{\sum n}$	$UCL_u=\bar{u}+3\sqrt{\dfrac{\bar{u}}{n}}$ $LCL_u=\bar{u}-3\sqrt{\dfrac{\bar{u}}{n}}$

① np = 不合格品数，c = 缺陷数，n = 每个子组的样本量，k = 子组数。

② 此公式用来设定控制界限。为避免计算麻烦，如果实际子组容量在平均子组容量20%波动范围内，可将平均子组容量作为共同的固定子组容量。

③ 如果控制下限为负数，用0代替。

表16.3 常数表

子组容量	均值—极差控制图			均值—标准差控制图			
	A_2	D_3	D_4	A_3	B_3	B_4	C_4
2	1.880	0	3.267	2.659	0	3.267	0.7979
3	1.023	0	2.574	1.954	0	2.568	0.8862
4	0.729	0	2.282	1.628	0	2.266	0.9213
5	0.577	0	2.114	1.427	0	2.089	0.9400
6	0.483	0	2.004	1.287	0.030	1.970	0.9515
7	0.419	0.076	1.924	1.182	0.118	1.882	0.9594
8	0.373	0.136	1.864	1.099	0.184	1.815	0.9650
9	0.337	0.184	1.816	1.032	0.239	1.761	0.9693
10	0.308	0.223	1.777	0.975	0.284	1.716	0.9727

表 16.4　常数表

子组容量	中位数—极差控制图			单值—移动极差控制图			
	A_2	D_3	D_4	E_2	D_3	D_4	d_2
2	--------	0	3.267	2.659	0	3.267	1.128
3	1.187	0	2.574	1.954	0	2.568	1.693
4	--------	0	2.282	1.628	0	2.266	2.059
5	0.691	0	2.114	1.427	0	2.089	2.326
6	--------	0	2.004	1.287	0	1.970	2.534
7	0.509	0.076	1.924	1.182	0.076	1.882	2.704
8	--------	0.136	1.864	1.099	0.136	1.815	2.847
9	0.412	0.184	1.816	1.032	0.184	1.761	2.970
10	--------	0.223	1.777	0.975	0.223	1.716	3.078

在解释控制图时，重要的一点是确定过程均值（中心线）相对于规格或目标的位置有没有发生变化。如果过程均值不在它应该在的位置处，要么是过程发生了改变，要么是规格发生了改变。为了分辨普通原因和特殊原因，必须对与控制界限有关的数据进行分析。上下控制界限并非就是规格界限，因此，并不能根据它对过程做出价值判定（好、差、临界状态）。

图 16.9 和图 16.10 给出了数据分析的步骤。为了让控制图能够作为一个监控工具

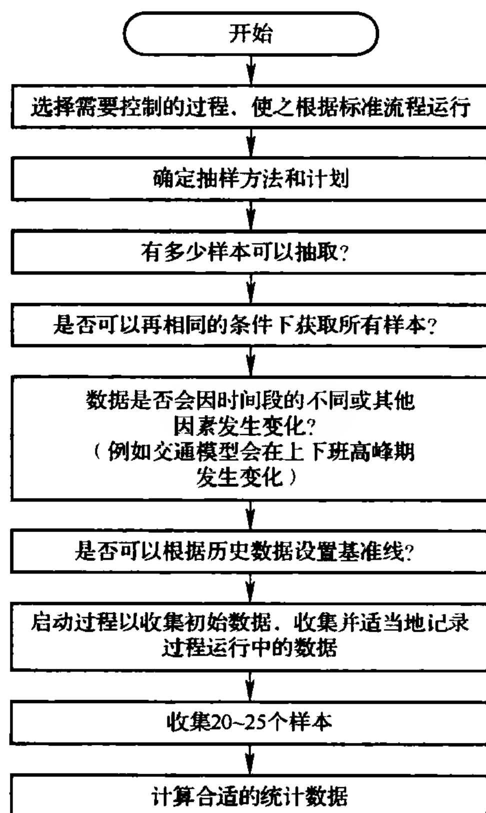


图 16.9　建立控制图的十个步骤

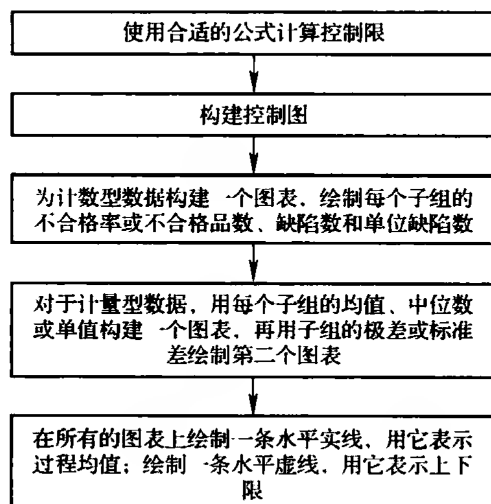


图 16.10　控制图的数据分析过程

使用，必须消除所有的特殊原因。当特殊原因再度出现时，控制图就会把它显示出来。普通原因是由类似于测量变差的统计误差引起的，受随机原因影响的零件通常都落在控制界限内；特殊原因是由过程失效引起的，例如设备故障。受其影响的零件通常都会落在控制界限外或呈现一种异常的趋势，如所有位于上控制限（UCL）曲线上方的点。图16.11所示为过程失效的判定过程。没有特殊原因的过程就是一个统计受控的过程。统计受控意味着过程是一致的；与此同时，必须检查过程是否符合规格界限。

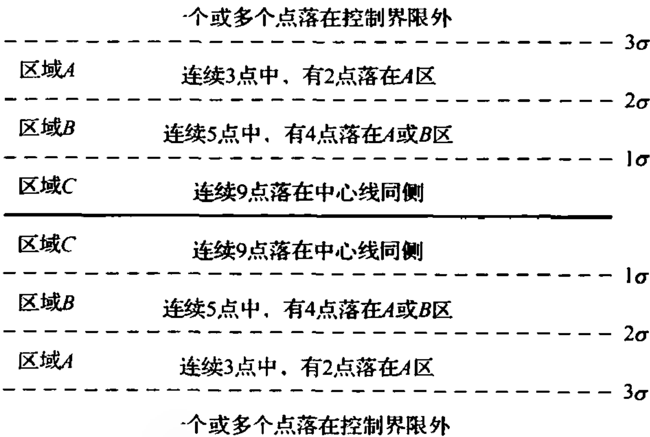


图16.11　过程失效的判定过程

在探究到特殊原因后，就必须要通过改变过程来消除它们；常规原因是产品所固有的，尝试改变它们，很可能会使局面更糟糕。只要过程不发生变化，控制界限就不会发生变化。表16.5介绍了7种常用的判定过程失控的规则。这些规则的图形化表示如图16.12所示。

表16.5　过程失控的判定规则

1	超过一个点落在控制界限外
2	连续3个点中，有2个落在A区
3	连续5个点中，有4个落在A区或B区
4	连续9个点落在中心线同侧
5	连续6个点递增或递减
6	连续14个上下交替分布
7	连续15个点落在C区

在确定一个过程失控之后，必须采取一系列的措施使其回归正常。表16.6介绍了一些改善措施的例子。对于表中任何答案为“是”的项，改进团队都应该将其视为潜在的特殊原因。

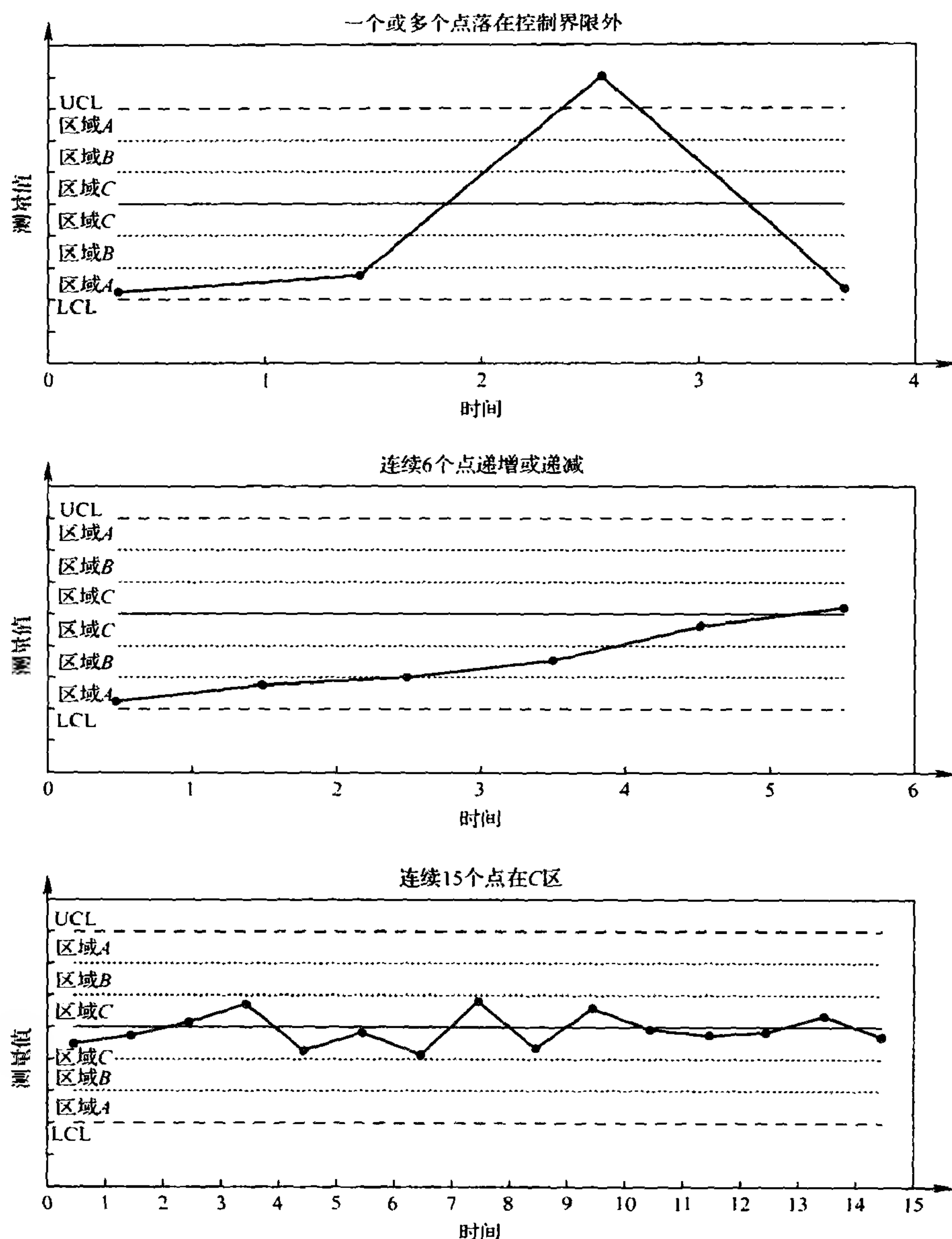


图 16.12　由表 16.5 生成的失控过程

表 16.6　过程分析中的常见问题

1	测量设备（方法）的精度是否存在差异？
2	不同的人员是否采用不同的方法？
3	过程是否会受到环境的影响（例如温度和湿度）？
4	测量环境是否已发生了巨大变化？

（续）

5	过程是否会受到可预测因素的影响（如工具的磨损）？
6	同一时间内是否有不同的人员参与到过程中？
7	输入到过程的内容是否已发生改变（例如原材料和其他信息）？
8	过程是否会受到员工疲劳的影响？
9	相关程序和策略是否已发生改变（如维修程序）？
10	过程是否经常调整？
11	样本是否来自不同过程的零件？是否来自不同的工作轮班？是否由不同的生产人员制造？
12	员工是否惧怕报告“坏消息”？

16.5 控制图案例

本节提供一些不同类型控制图的案例，包括：

X（bar）图——显示测量均值的波动；

r 图——显示测量值极差的波动；

c 图——显示不合格数的波动；

u 图——显示单位产品上不合格数的波动；

p 图——显示不合格率的波动；

np 图——显示不合格品数的波动。

案例 16.2

用 *X*（bar）图和 *r* 图分析某机械加工厂生产的零件的重量。该工厂会在 20 个不同的时间点分别进行取样，每次取 5 个产品并测量其重量（数据见表 16.7）。

表 16.7 某机械加工厂生产的零件的数据

子组号	*A*	*B*	*C*	*D*	*E*	*X*	*R*
1	1.4	1.2	1.3	1.4	1.2		
2	1.3	1.2	1.3	1.5	1.3		
3	2.7	1.3	1.4	1.2	1.2		
4	1.4	1.2	1.3	1.3	1.4		
5	1.5	1.1	1.7	1.3	1.3		
6	1.8	1.2	1.5	1.5	1.4		
7	1.5	1.2	1.3	1.3	1.2		
8	1.7	1.7	1.2	1.2	1.1		
9	1.8	1.8	1.7	1.8	1.5		
10	1.1	1.2	1.8	1.6	1.3		
11	1.2	1.3	1.4	1.4	1.4		
12	1.3	1.9	1.9	1.5	1.5		
13	1.4	1.8	1.7	1.1	1.3		
14	1.8	1.9	1.5	1.4	1.4		
15	1.1	1.3	1.1	1.8	1.5		

（续）

子　组　号	A	B	C	D	E	X	R
16	1.8	1.9	1.7	1.6	1.3		
17	1.2	1.4	1.3	1.2	1.4		
18	1.1	1.1	1.7	1.2	1.3		
19	1.8	1.6	1.5	1.7	1.8		
20	1.1	1.3	1.3	1.4	1.3		

解：因为子组容量 $n=5$，因此选择 X(bar) 图和 r 图进行控制。

① 首先计算每个子组的均值和极差：

均值 $\overline{X}$ = 每个子组内样本的总和/样本容量；

对于子组 1，$\overline{X}=(1.4+1.2+1.3+1.4+1.2)/5=1.3$；

极差(R) = 子组内最大值 - 最小值；

$R_1=(1.4-1.2)=0.2$。

计算 $\overline{X}$ 和 R 列的总和。

② 平均均值和平均极差计算：

总均值($\overline{\overline{X}}$) = 所有子组均值的总和/子组数 = 28.54/20 = 1.43；它也叫总平均值(Grand Average)，用于设定控制图的中心线。

所有子组极差的均值（$\overline{R}$）= 总均值/子组数 = 9.0/20 = 0.45，它用于设定极差图的中心线（见表 16.8）。

表 16.8　用于计算，绘图的数据

子　组　号	A	B	C	D	E	X	R
1	1.4	1.2	1.3	1.4	1.2	1.3	0.2
2	1.3	1.2	1.3	1.5	1.3	1.3	0.3
3	2.7	1.3	1.4	1.2	1.2	1.3	0.5
4	1.4	1.2	1.3	1.3	1.4	1.3	0.2
5	1.5	1.1	1.7	1.3	1.3	1.3	0.6
6	1.8	1.2	1.5	1.5	1.4	1.4	0.6
7	1.5	1.2	1.3	1.3	1.2	1.3	0.3
8	1.7	1.7	1.2	1.2	1.1	1.3	0.3
9	1.8	1.8	1.7	1.8	1.5	1.7	0.3
10	1.1	1.2	1.8	1.6	1.3	1.4	0.7
11	1.2	1.3	1.4	1.4	1.4	1.3	0.2
12	1.3	1.9	1.9	1.5	1.5	1.6	0.6
13	1.4	1.8	1.7	1.1	1.3	1.4	0.7
14	1.8	1.9	1.5	1.4	1.4	1.6	0.5
15	1.1	1.3	1.1	1.8	1.5	1.3	0.7
16	1.8	1.9	1.7	1.6	1.3	1.6	0.6
17	1.2	1.4	1.3	1.2	1.4	1.3	0.2
18	1.1	1.1	1.7	1.2	1.3	1.2	0.6
19	1.8	1.6	1.5	1.7	1.8	1.6	0.3
20	1.1	1.3	1.3	1.4	1.3	1.2	0.3
						28.0	9.0

③ 控制限计算：

$$UCL_X = \overline{\overline{X}} + A_2\overline{R} = 1.43 + (0.577 \times 0.45) = 1.69$$

$$LCL_X = \overline{\overline{X}} - A_2\overline{R} = 1.43 - (0.577 \times 0.45) = 1.17$$

即大约99.73%（3σ限）的样本均值落在1.17～1.69的范围内；

$$UCL_R = D_4\overline{R} = 2.114 \times 0.45 = 0.951$$

$$LCL_R = D_3\overline{R} = 0 \times 0.45 = 0$$

即大约99.73%（3σ限）的样本极差落在0～0.951的范围内。

案例16.3

某机械加工厂生产的零件重量如表16.9所示。

表16.9　案例16.3的数据

观测序号	样本测量值（X）	MR
1	1.4	
2	1.3	
3	1.7	
4	1.4	
5	1.5	
6	1.8	
7	1.5	
8	1.7	
9	1.8	
10	1.1	
11	1.2	
12	1.3	
13	1.4	
14	1.8	
15	1.1	
16	1.8	
17	1.2	
18	1	
19	1.8	
20	1.1	
总计	28.9	

由于生产速度的原因，在一个观测的时间段内只能抽取一个零件来估计其重量。对于这种情况，可以使用移动极差图（Moving Range Chart）来分析重量的变化。

解：由于每次只有一个数据，我们选择移动极差图。

① 首先计算 MR：

$MR = |R_n - R_{n-1}|$ = 多个连续观测值差的绝对值，也叫二样本移动极差（最常见的移动极差）。

第一个数据没有极差。

第一个移动极差 $MR_1 = |1.4 - 1.3| = 0.1$。

计算表 16.10 中所有样本观测值的总和和所有 MR 的总和。

表 16.10　用于计算，绘图的数据

观测序号	样本测量值（X）	MR
1	1.4	N/A
2	1.3	0.1
3	1.7	0.4
4	1.4	0.3
5	1.5	0.1
6	1.8	0.3
7	1.5	0.3
8	1.7	0.2
9	1.8	0.1
10	1.1	0.7
11	1.2	0.1
12	1.3	0.1
13	1.4	0.1
14	1.8	0.4
15	1.1	0.7
16	1.8	0.7
17	1.2	0.6
18	1	0.2
19	1.8	0.8
20	1.1	0.7
总计	28.9	6.9

② 然后计算所有均值的平均值和子组极差的平均值：

总均值（$\overline{X}$）= 所有观测值总和/观测数 = 28.9/20 = 1.45。

$\overline{X}$ 也称为总平均值，它用来确定均值图的中心线。

所有子组移动极差的平均值 $M\overline{R}$ = MR 的总和/极差数 = 6.9/19 = 0.36，$M\overline{R}$用作移动极差图的中心线（平均值）。

③ 最后决定控制限：

$$UCL_X = \overline{X} + (E_2 \times M\overline{R}) = 1.45 + (2.659 \times 0.36) = 2.41$$

$$LCL_X = \overline{X} - (E_2 \times M\overline{R}) = 1.45 - (2.659 \times 0.36) = 0.49$$

$$UCL_{MR} = D_4 \times M\overline{R} = 3.267 \times 0.36 = 1.18$$

$$LCL_{MR} = D_3 \times M\overline{R} = 0 \times 0.36 = 0$$

绘出的均值图如图 16.13 所示，极差图如图 16.14 所示。

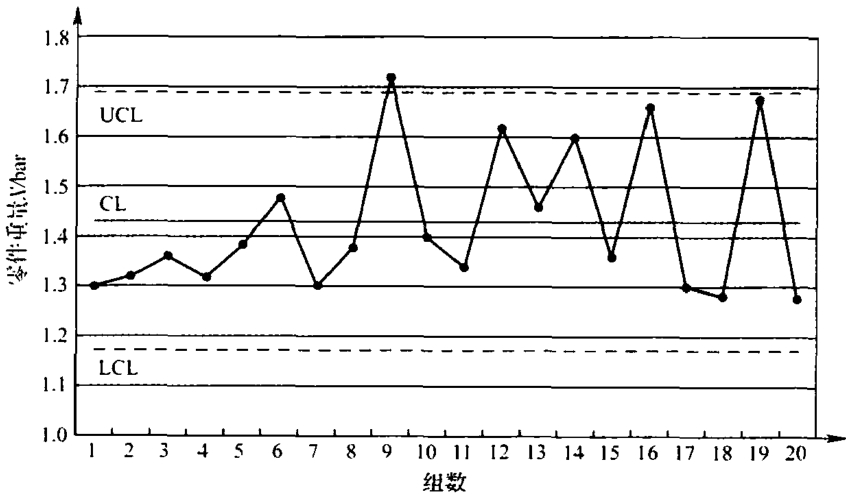


图 16.13 均值图

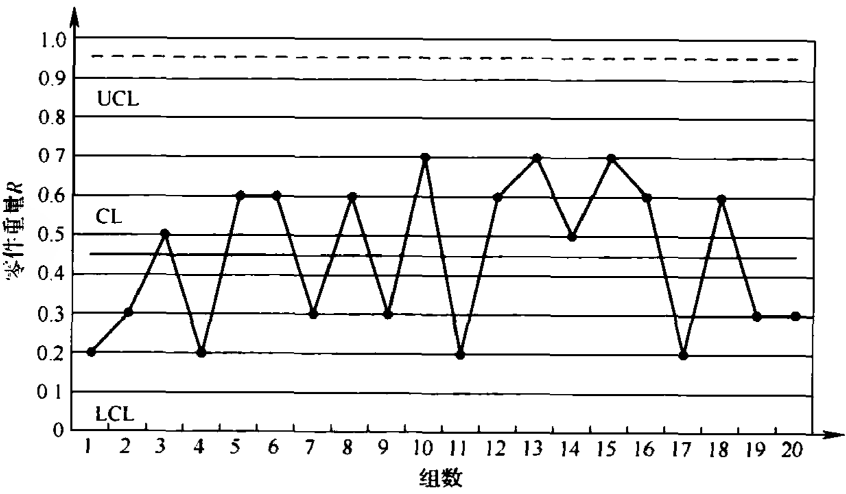


图 16.14 极差图

注意：在这个案例中，用来计算系数 E_2、D_2 和 D_4 值的样本数为 2，因为此案例中我们采用的是二样本极差。如果使用三样本极差，极差的数量将减少到 18，所使用的常数值也将相应地发生变化。

案例 16.4

根据下面信息分析某机械加工厂生产零件的重量。在样本容量为 50，统计 10 周内

的不合格品数（相关数据见表 16. 11）。

表 16. 11　案例 16. 4 的数据

周　次	缺 陷 数
1	9
2	7
3	4
4	2
5	4
6	5
7	2
8	3
9	5
10	5
总计	46

解：由于此案例中的样本量恒定，它和不合格品数均为计数型数据，故选择 *np* 控制图。

首先确定均值：

平均不合格率 $\bar{p}$ = 所有不合格品数/所有样品数，即

$$\bar{p}=\frac{46}{(n)(\text{weeks})}=\frac{46}{(50)(10)}=0.092$$

不合格品数均值 $n\bar{p}=50\times0.092=4.6$ 或 $n\bar{p}=46/10=4.6$。

然后确定控制限：

$$\text{UCL}=n\bar{p}+3\sqrt{n\bar{p}(1-\bar{p})}=4.6+3\sqrt{4.6(1-0.092)}=10.731$$

$$\text{LCL}=n\bar{p}-3\sqrt{n\bar{p}(1-\bar{p})}=4.6-3\sqrt{4.6(1-0.092)}=0$$

注意：下控制限（LCL）小于 0，用 0 代替。

绘出的单值图和移动极差图分别如图 16. 15 和图 16. 16 所示，绘制的 *np* 控制图如图 16. 17 所示。

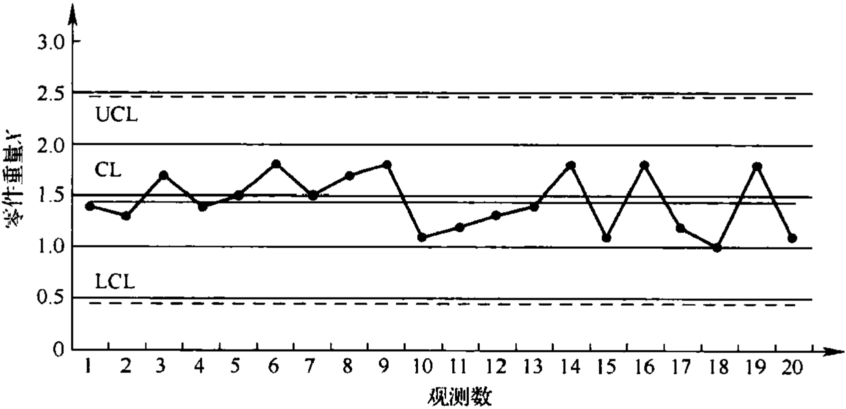


图 16. 15　单值图

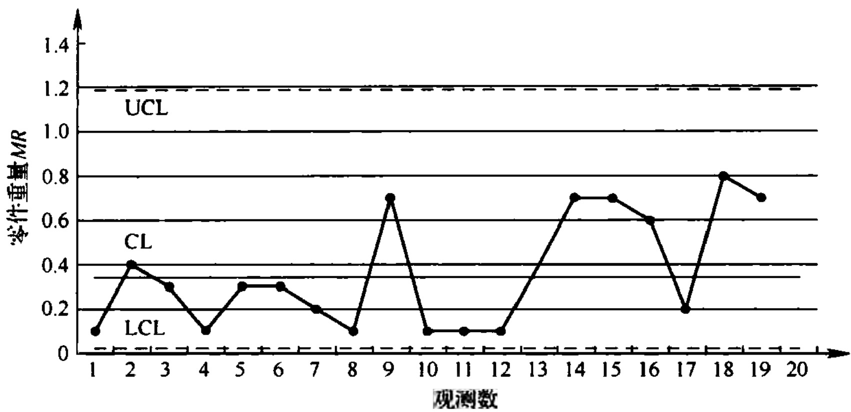


图 16.16　移动极差图

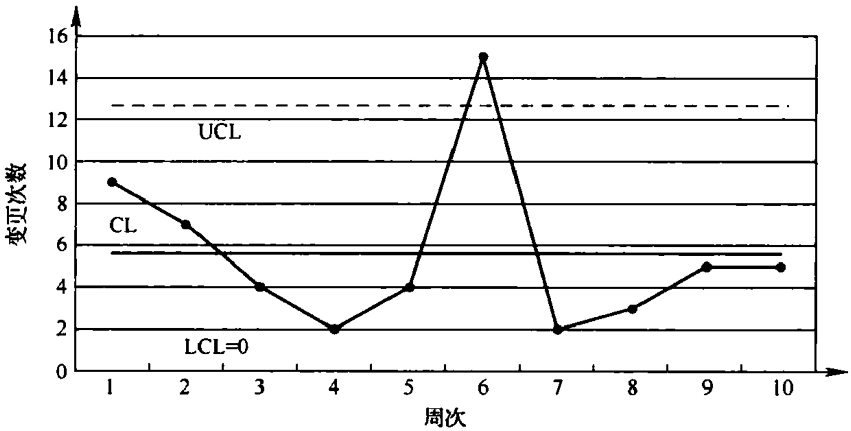


图 16.17　不合格品数控制图（*np* 图）

案例 16.5

某公司统计其产品规格由于工程变更或应合同管理部门要求发生的变更次数，统计周期为 10 周，且变更次数总和超过 50。

解：由于变更次数为离散数据，样本容量恒定（每周），而且变更次数用缺陷数来表述，因此用 c 控制图最为合适（见表 16.12）。

首先确定中心线（$\overline{C}$）：

$\overline{C}$ = 所有缺陷的总和/所有子组数 = 56/10 = 5.6（每周变更次数）

其次确定控制限（如果下控制限 <0，则以 0 代替）：

$$\mathrm{UCL} = \bar{c} + 3\sqrt{\bar{c}} = 5.6 + 3\sqrt{5.6} = 12.699$$

$$\mathrm{LCL} = \bar{c} - 3\sqrt{\bar{c}} = 5.6 - 3\sqrt{5.6} = 0$$

最后绘制控制图。

表 16.12 案例 16.5 的数据

周 次	缺 陷 数
1	9
2	7
3	4
4	2
5	4
6	15
7	2
8	3
9	5
10	5
Total	56

参考文献

Ackermann C. S. , and J. M. Fabia. 1993. Monitoring supplier quality at PPM levels. IEEE Transactions on Semiconductor Ma nufacturing 6 (2): 189-195.

JEDEC. 1995. Standard JESD16 – A. Assessment of average outgoing quality levels in parts per million (PPM) . Electronic Industries Association, Alexandria, VA.

练习

16.1 对于表 16.13 中给定的数据组，请判断使用何种控制图能够最合适地描绘这些数据，并阐述选择的理由，可选控制图类型包括：*c* 图、*u* 图、*p* 图、*np* 图、*X*(bar) -r 图或者 $X\text{-}R_m$ 图。

表 16.13 练习 16.1 的数据组

数据详情	控制图类型
最近 5 周连续每周抽取相同数量的样本进行检验：第一周发现 10 个不合格品，第二周发现 8 个，第三周发现 6 个，第四周发现 9 个，第五周发现 7 个	
最近 4 周每周抽取不同数量的样本（40 到 60 个之间）进行检验：第一周发现的单位缺陷数是 1.2，第二周是 1.5，第三周是 1，第四周是 0.8	
每天测量 10 个产品的厚度，测量 1 周	
最近 4 周连续每周抽取相同数量的样品进行检验：第一周发现 8 个缺陷，第二周 12 个，第三周 10 个，第四周 9 个	
每天测量一个产品的厚度，测量 7 天	
最近 3 周连续每周对过程的不合格率进行监控：第一周的产品不良率为 10%，第二周为 20%，第三周为 15%	

16.2 对电镀槽中铜的含量每天检测3次，结果以百万分之几（ppm）表示。检测10天的结果用均值—极差的形式记录如表16.14：

表16.14 练习16.2的数据

天　数	均　值	极　差
1	5.45	1.21
2	5.39	0.95
3	6.85	1.43
4	6.74	1.29
5	5.83	1.35
6	7.22	0.88
7	6.39	0.92
8	6.5	1.13
9	7.15	1.25
10	5.92	1.05

（a）确定上、下控制限。

（b）过程是否统计受控？

（c）假设规格是6.0±1.0，估计过程的C_p和C_{pk}，工艺能力是否充分？

16.3 通过自动和人工组合的方式组装印制电路板。在此过程中，回流焊用机构/电子连接器把含铅零件焊接到电路板上。电路板连续地通过焊接过程，且每个小时抽取5片板进行检查。每块板子上的缺陷都会被记录下来。表16.15展示了其中的20组数据。请问哪种控制图最适合？为什么？计算控制限并绘制控制图？过程是否受控？是否需要改进？

表16.15 练习16.3的数据

样　本	缺　陷　数	样　本	缺　陷　数
1	6	11	9
2	4	12	15
3	8	13	8
4	10	14	10
5	9	15	8
6	12	16	2
7	16	17	7
8	2	18	1
9	3	19	7
10	10	20	13

16.4 12个同样的产品进行1000h的测试，其中7个分别在250h、450h、510h、625h、750h、825h和979h出现失效。所有失效的产品都被取走，且无新的产品进行替换。请计算90%置信水平下，MTBF的单侧置信上下限？和90%置信水平下，200h内可靠度的双侧置信界限。

16.5 某根轴的直径的标称规格是60±3mm，每小时测量6次并予以记录。连续测

量 8 小时，每小时的均值和极差值如表 16.16 所示：

表 16.16 练习 16.5 的数据

小　时	均　值	极　差
1	62.54	1.95
2	60.23	2.03
3	58.46	1.43
4	59.95	1.29
5	61.58	0.78
6	57.93	1.48
7	61.56	0.86
8	57.34	1.35

（a）确定上下控制限。

（b）确定过程是否统计受控？

（c）估计过程的 C_p 和 C_{pk} 值，并确定工艺能力是否充足？

16.6　轴的直径规格是 212 ±2mm。每次测量 1 根轴，先后测量 30 根轴，数据如表 16.17 所示（单位 mm）：

表 16.17 练习 16.6 的数据

212.1①	214.2	213.7	212.7	212.5	212.7②
212.8	213.0	212.9	212.3	212.5	212.1
211.8	213.5	212.0	213.0	214.5	212.3
212.2	211.9	213.2	212.7	211.9	212.3
212.0	212.8	213.9	212.6	214.0	212.4③

① 第 1 个测量值；

② 第 6 个测量值；

③ 第 30 个测量值。

（a）确定长度为 3 的单值—移动极差图的控制限。

（b）从控制图确定该过程是否受控？

（c）确定该过程的能力指数（C_p 和 C_{pk}）。

（d）确定该过程的不合格品率。

国际视野　科技前沿

ISBN 978-7-111-34930-3
封面设计：鞠杨
定价：79.00元

地址：北京市百万庄大街22号
电话服务
社服务中心：(010)88361066
销售一部：(010)68326294
销售二部：(010)88379649
读者购书热线：(010)88379203

邮政编码：100037
网络服务
门户网：http://www.cmpbook.com
教材网：http://www.cmpedu.com
封面无防伪标均为盗版

上架指导：工业技术／机械
／可靠性、故障诊断

